

FACILITY FORM 808

N66 271 98	
(ACCESSION NUMBER)	(THRU)
225	1
(PAGES)	(CODE)
CR-57011	10
(NASA CR OR TMX OR AD NUMBER)	(CATEGORY)

NASA CR-57011

GPO PRICE \$
CFSTI PRICE(S) \$
Hard copy (HC) 6.00
Microfiche (MF) 1.25-

11 853 July 85

RESEARCH ON FAILURE FREE SYSTEMS
WITH SUPPLEMENTAL INFORMATION

December, 1963

Distribution of this report is provided in the interest of information exchange and should not be construed as endorsement by NASA of the material presented. Responsibility for the contents resides in the author or organization that prepared it.

[REDACTED]

Prepared under Contract No. NASw-572 by
THE WESTINGHOUSE ELECTRIC CORPORATION
Baltimore, Maryland

for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

[REDACTED]

LIST OF CODE LETTERS AND COMPANY NAMES

Company A--Westinghouse Electric Corporation

Company B--Signetics Corporation

Company C-- Curtis Instruments, Inc.

Company D--Fairchild Camera & Instrument Corporation

Company E--Sylvania Electric Products, Inc.

Company F--Siliconix Incorporated

Company G--Texas Instruments Incorporated

Company H--Motorola, Inc.

Company I--Amp, Inc.

LIST OF CODE NUMBERS AND TRADE NAMES

Device 1--Memistor

Device 2--Amp-Mad

CASE FILE COPY

TABLE OF CONTENTS

	<u>Page</u>
PURPOSE	1
SUMMARY	3
CONCLUSIONS AND RECOMMENDATIONS	9
Appendix 1 - Design and Testing of Redundant Systems	
Appendix 2 - Reliability of Imperfect Redundant Systems	
Appendix 3 - A Survey of Components for Adaptive Restoring Circuits	
Appendix 4 - Transor Analysis	
Appendix 5 - Comparison of Dynamic and Threshold Restorers	
Appendix 6 - Self Repair Techniques	

PURPOSE

This final report is prepared in accordance with the requirements of Contract NASw-572, "Research on Failure Free Systems", between the National Aeronautics and Space Administration and the Westinghouse Electric Corporation (reference WGD-38521). The research that is reported herein has the general objective of the advancement of the state-of-the-art in the design of highly reliable electronic systems associated with the national space effort. The design objectives which are studied are those which permit the proper operation of systems to be relatively independent of the effects of individual component or module failures within systems. The scope of this objective includes the use of the more conventional techniques of multiple-line, majority voted redundancy, as well as the study of self-repair and advanced voting techniques. The research has been divided into the following major tasks:

TASK 1: IMPLEMENTATION

TASK 2: ADVANCED VOTING TECHNIQUES

TASK 3: SELF REPAIR TECHNIQUES

SUMMARY

TASK 1 - IMPLEMENTATION.

This portion of the study is concerned with developing suitable circuits, systems, and testing techniques for use with currently available redundancy techniques. The circuit and system design is expected to be suitable for general use in spaceborne or ground support equipment, free from extremely detrimental failure modes, and compatible with whatever testing techniques are to be applied. The testing techniques are expected to be suitable for a wide variety of applications. They are, therefore, similarly varied according to the purpose of the testing, the system configuration involved, and the information which is available for the test. The testing of redundant systems represents a unique problem, since individual component or module failures do not indicate their occurrence by affecting the system performance. The various purposes for testing are indicated by the following types of diagnostic tests which have been considered:

The verification that all signal-processing elements are working properly, or additionally that the voters are capable of transmitting a correct signal, or further, that all signal processors and voters work properly under all possible design conditions. This may be further extended to include the verification that any additional hardware which is added for the testing is also capable of proper operation. This range of test requirements is also encountered when the purpose of the tests is not only to detect any failures, but to locate these failures to facilitate repair or replacement in redundant systems where repair is desired, or systematic maintenance is used.

Another type of testing is referred to as "statistical measure of quality", which obtains a limited amount of information concerning the failure pattern existing within the system to estimate the reliability of the system. Many different types of tests can be used to obtain this information, depending on the confidence required for the reliability estimate, the cost of obtaining the information, and the type of analysis which will be applied to that information.

Much of the preliminary work necessary for the determination of suitable circuitry for redundant systems has been described in an earlier Company A report, "Failure Effects in Redundant Systems"¹. The report describes in detail the effects of catastrophic component

1. A. R. Helland, W. C. Mann, "Failure Effects in Redundant Systems", Westinghouse Report EE 3351, March 1963.

failures which were induced into a laboratory model of a portion of a typical redundant system. Potentially serious detrimental failures which might occur are discussed. A major portion of the report is concerned with the random failure simulations and their results. Briefly, a computer program generates random failure lists using available reliability data for each part. Each failure list includes all component failures which might have occurred in a typical system which had been operated for the specified time interval, and therefore simulates the actual testing of such a system. The indicated failures are induced into the system, which is then tested to determine whether it is capable of performing all of its design functions, or if it has failed. This actual test result can be compared with the analytical result which would have been obtained with the same group of failures, to test the validity of the assumptions used for the analytical result. These tests showed that the most common analytical model is excessively pessimistic for a well-designed system. For these tests it predicted more system failures than actually occurred by a ratio of more than 2:1. The reasons for this departure, and more accurate analytical models, are discussed. A new technique is described which permits the reliability of a redundant system to be estimated by the product of exponentials, using the failure rates of the components or modules involved. Finally, several circuit design considerations are discussed.

The results of implementation studies as part of the research on failure free systems have been previously published in special technical reports. Two major areas of interest are discussed in Special Technical Report No. 3, "Circuits and Circuit Testing for Spaceborne Redundant Digital Systems". The entire report is reproduced as Appendix 1 of this final report. The first portion of the report is concerned with efficient initial design and contains a discussion of several possible circuit implementations. The latter portion is concerned with the diagnostic testing of a multiple line, majority logic redundant system. Several techniques are described for detecting and locating failures within an operating redundant system to greatly increase reliability. The report is summarized below.

Section I contains a discussion of the general problems concerned with the design and testing of redundant systems. These problems include the most appropriate choice of circuit implementation, special design requirements, and the realization of high system reliability with available circuits.

Section II contains a discussion of the possible use of magnetics to reduce the total power consumption and provide non-volatile storage in redundant spaceborne systems. Magnetics appear to be most useful for applications requiring memory associated directly with simple forms of logic, or for non-volatile data storage when the data is altered at very slow rates, but is not recommended for general logic use.

Section III contains descriptions and comparisons of types of semiconductor circuits suitable for use in redundant systems. Since integrated circuits offer many important advantages for redundant systems, they are chosen as a basis for system design with semiconductor circuitry. Since custom design of integrated circuits is not especially practical for low volume operation, the circuit design problem includes the choice of the most suitable type of available circuits. Integrated Diode-Transistor Logic elements were chosen as the most appropriate for general use. A majority voter restoring element, which is not subject to the detrimental failure modes found to be characteristic of conventional elements, is designed using positive logic D-TL NAND elements.

The discussion of Section IV is concerned with the testing of redundant systems. Various solutions to the problem of failure detection within a redundant system are discussed in this section; some are more suitable for simple failure detection, others also provide information concerning the location of any failures. The failure detection tests alone are expected to be most suitable for initial acceptance and verification tests to indicate that all parts are working. The combined detection and location techniques are most applicable to systems where additional information is required to facilitate repair or replacement of individual parts of the system.

It is shown that failure location and maintenance of a redundant system does not require the test equipment and operator skill which are usually required to maintain a conventional non-redundant system. Techniques are described which permit a redundant system to be systematically maintained to provide much higher operational reliability than possible without maintenance. It is shown that a major portion of the maintenance may be performed during normal system operation.

The partial testing of imperfect redundant systems to estimate future reliability is discussed in part two of Special Technical Report No. 4, "Transor Decision Functions and Statistical Measure of Quality". The second part of the report is reproduced as Appendix 2 of this final report.

The objective of this portion of the study has been to develop a test philosophy from which a good statistical estimate of the probability of mission success could be made from a limited amount of test data. Several possibilities have been formulated. The failure masking characteristics of redundant systems prohibit the use of simple test programs which merely determine the performance capability of the system at the time of test. Such programs

cannot differentiate between systems containing many component failures with correspondingly many stages vulnerable to succeeding failures, or few component failures with few vulnerable stages. Because the probability of mission success after the time of test is heavily influenced by the component failure pattern existing at the time of test, a test program must be devised from which mission reliability can be predicted with a reasonably high degree of confidence. The general complexity and microminiature size of modern systems generally precludes the possibility of testing each signal processor in each stage.

In the proposed extension of this study the various philosophies will be considered in more detail, and an effort will be made to evaluate the usefulness of each one with the purpose of determining which of the candidate philosophies provides the most accurate estimate of probability of mission success for a fixed cost of testing.

TASK 2 - ADVANCED VOTING TECHNIQUES

This study is concerned with advancing the state-of-the-art in developing new restoring circuits for use in redundant systems. Several advanced voting techniques have been studied as part of the research on failure free systems. The results of the Adaline-Neuron study and the initial results of the Transor study have been previously published as special technical reports. Further study of Transor and a new dynamic restorer (the Hamming Distance Restoring Circuit) has been conducted, but the results have not been previously published. These results are, however, contained in Appendix 5 of this report.

The results of the study of the Adaline-Neuron adaptive voter with continuously variable input weighting have been previously published as Special Technical Report Number 1, "A Survey of Adaptive Components for Use in Failure Free Systems". It is reproduced as Appendix 3 of this report. Briefly, it concludes that suitable analog memory devices are not currently available for use in this class of adaptive voters, although the mercury cell integrator with photoelectric readout is apparently the most suitable technique.

Since the Adaline-Neuron adaptive voter requires an analog memory for each input, the selection of a suitable input device is important to realize a practical adaptive voter. Several types of analog memory devices were surveyed in order to evaluate their suitability for use in implementing an adaptive voter for redundant systems. It is desirable that the devices be simple, reliable, relatively linear, and store the analog variable weighting for a relatively long time. It was found that most of the available devices which have been developed for pattern recognition or learning machines are too complex, unstable, or unreliable for use in adaptive voters.

Devices which were included in the survey included the Device 1 plated resistor, the solion iodine ion cell, the mercury cell integrator (with either capacitive or photoconductive readout), the MAD magnetic integrator, the orthogonal core integrator, the second harmonic magnetic integrator, and the magnetostrictive integrator. The mercury cell integrator with photoconductive readout appears to be the most suitable device among those which were surveyed. It incorporates an electroplating technique for providing the continuously variable input weighting for adaptive voters, with relatively good stability, reversibility, and permanent storage. Since it is a four terminal device with electrical current as the input and electrical resistance as the output, it is relatively simple and generally compatible with conventional circuitry. It is, however, currently in a relatively early state of its development as a device for general use. It appeared that any detailed circuit design for adaptive voters should not be undertaken before the expected progress in the development of more effective cells is accomplished.

The proposed continuation of the development of this class of adaptive voters includes monitoring the state of the art in the development of more effective devices, followed by the design and breadboard construction of at least one Adaline-Neuron adaptive restorer, or preferably a small redundant subsystem using these restorers, in order to demonstrate their effectiveness in redundant systems.

The objective of the Transor study portion of the research was to evaluate the Transor Restoring Circuit for possible use as a replacement for threshold voters in redundant systems. In the process of performing this evaluation, another dynamic restorer, the Hamming Distance Restoring Circuit, was invented. The study was extended to include an evaluation of both circuits.

The initial portion of this study has been reported in part one of Special Technical Report No. 4, "Transor Decision Functions and Statistical Measure of Quality" which is reproduced as Appendix 4 of this final report. In that report, analytical reliability expressions for systems using Transor restorers are obtained for the case when signal processors are restrained by certain failure mode assumptions. An appendix to that report shows how the probability of occurrence of various failure modes might be computed. The results of later portions of this work are presented in Appendix 5 of this final report. In these results, general reliability expressions for the Transor and the Hamming Distance Restoring Circuit are obtained which are relatively free of restrictive assumptions. A computer simulation program which was developed for use in the evaluation, is described and some results obtained from the program are discussed. Finally, the conclusion is drawn that the Hamming Distance Restoring Circuit is always superior to the Transor but that it is as good as or better than the threshold voter only in certain failure mode environments.

TASK 3 - SELF REPAIR TECHNIQUES.

This study is concerned with the development of new, more efficient means for employing redundant equipment. Using these techniques, a system may be designed to absorb more internal failures without system failure than is possible with the same amount of fixed, multiple-line redundant equipment. The results of this study have been previously published as Special Technical Report No. 2, "Self Repair Techniques for Failure Free Systems". The report is reproduced as Appendix 6 of this final report.

As a part of the effort to develop hyper-reliable systems, Company A has devised a class of techniques for using redundant blocks of circuitry more effectively than has been done previously. The systems using these techniques are similar to the familiar multiple-line, majority-voted redundant systems except blocks of circuitry are allowed to shift around as component failures leave certain subsystem functions more vulnerable than others to succeeding failures. The object of this phase of the study has been to devise several general patterns in which systems could be organized to absorb relatively large numbers of internal failures without system failure and to develop a means for evaluating the effectiveness of the various patterns for performing this function.

Three broad classes of organization patterns have been developed, and several specific patterns within each class have been examined. A versatile computer simulation program has been written from which approximate reliability vs. time curves and a variety of other pertinent information about each pattern can be directly obtained. Both of the patterns which have been developed and the computer program have been described in detail in Appendix 6.

A three-part program has been proposed for future study in this area. In the first part, the computer simulation program will be used as an evaluation tool for establishing a set of rules for designing optimum or near-optimum self-repairing systems. The rules will be primarily concerned with the organizational patterns to be used and with the maximum allowable ratio of repair circuitry complexity to signal processor complexity. Secondly, an implementation study has been proposed to determine effective means for implementing the organization patterns which have been and will be devised. Finally, an appropriate study vehicle will be selected and designed with sufficient detail that a breadboard model could be constructed from the specifications produced. Such a vehicle design is required in order to verify the usefulness of both the organizational pattern theories and the implementation techniques which are being developed.

CONCLUSIONS AND RECOMMENDATIONS

TASK 1 - IMPLEMENTATION

1. Design of Redundant Systems

Redundancy is a powerful tool for achieving extended reliability, but effective design is required to achieve the reliability goals with a minimum of additional complexity. Although magnetic logic is often cited as having several advantages applicable to spaceborne computers, the use of magnetic logic is limited to special applications. Magnetic logic is not particularly suited for general logic use in redundant systems, due to the lack of steady output signals, low speed capability, high peak power requirements, and the complexity required for general logic functions. It appears that no proven magnetic restoring element exists which is suitable for general use in redundant systems. Magnetic logic does, however, offer non-volatile storage and very low average power for slow speed operation. Magnetic devices appear to be suited to special applications where certain logic functions, such as transfer and OR, are intermixed with the memory function, and very low speed capability is acceptable. It is useful for low speed shift registers, counters, and timers which consume negligible standby power.

Integrated semiconductor circuitry offers many desirable characteristics for use in redundant spaceborne systems, including small size, reduced weight and power consumption and high frequency capability. A comparison of the currently available integrated logic elements indicates that diode-transistor logic (D-TL) is the most suitable for general logic use in redundant spaceborne systems. A majority voting restorer, designed using interconnected NAND elements, has been described which is not subject to the detrimental failures of more conventional restoring elements.

2. Testing of Redundant Systems

It is a characteristic of redundant systems that they offer a high reliability for a period of time after the initially failure free condition, and that the system reliability decreases rapidly when internal failures are present. It is therefore important to insure that no initial failures exist in a redundant system to obtain maximum system reliability. Since an initially failure free, order three system can withstand any single failure, as well as a relatively large number of randomly scattered failures, it offers very high reliability for the period of time when the probability of individual failures is low. Techniques are described which permit even higher reliability by the use of systematic maintenance of a redundant systems.

It has been shown that a relatively simple technique called singular rank testing may be used to determine that all of the replicated signal processors in a redundant system are working properly, and that the majority voters are sufficiently failure free to insure that the system is not vulnerable to single failures. The system is monitored to determine if each individual rank is able to perform all system functions correctly, in a manner similar to the verification of a non-redundant system. This testing places no restrictions on system size or configuration. A somewhat more complicated testing procedure, referred to as interwoven rank testing, has been described which will completely test all voters to insure that they will make correct decisions for all possible input combinations.

Although a redundant system is more complex than its conventional counterpart, failure location within a working system does not require the operator skill and simulation equipment usually required to locate failures in a non-redundant system. Since a working redundant system always has at least one correct signal available at each stage in the system, these correct signals may be used as a basis of comparison. A difference detector on the signal processor outputs to restorers may be used to indicate either permanent or sporadic failures among these signal processors. The failure location techniques described may be performed during normal operation, since they do not jeopardize system operations.

3. Reliability of Imperfect Redundant Systems

The mission reliability of an operating redundant system which contains internal failures depends strongly on the number and location of initial circuit failures, as well as the failure rates of the circuits which make up the system.

One very important task is the design of simple and efficient tests to be performed at the beginning of a mission. These tests are required to obtain the information required for the reliability estimates. A maximum amount of information is desired from a minimum number of tests. The work which has been done will provide a basis for future efforts in this area.

Several tests are proposed that may be made just before a mission is to begin to determine, at least approximately, the mission reliability without complete information on the state of the system. It proposes some procedures for using the results of the tests to estimate the mission reliability with varying degrees of accuracy. A procedure for making the decision on the useability of the system without estimating the mission reliability is also presented.

Although a basis for future study has been provided, the details of these procedures are still to be worked out and the accuracy of their results are still uncertain. It is recom-

mended that efforts be made to develop an appropriate measure for comparing the techniques so that they may be evaluated relative to a common scale.

TASK 2 - ADVANCED VOTING TECHNIQUES

1. Components for Adaptive Restorers

A survey has been conducted of several devices which are potentially suitable for use in the Adaline-Neuron adaptive voter. The survey concludes that none of the suggested devices were sufficiently developed to justify the immediate circuit implementation of an adaptive voter.

In general, magnetic devices do not appear to be suitable for use in adaptive voters, due to their environmental sensitivity and complexity required for useful operation. Similarly electro-chemical devices do not appear to have sufficient simplicity, stability and compatibility with electronic circuitry to justify their use in adaptive voters.

The mercury cell integrator with photoelectric readout appears in principle to offer the most attractive approach because of its simplicity, stability and general compatibility with conventional circuitry. Since the output is essentially a variable resistance proportional to the interval of the control input current, the device offers the possibility of providing a simple interface with standard circuitry. The mercury cell integrator is, however, still in a rather primitive state of development. It is recommended that detailed circuit design should not be undertaken until further device development is completed and that present effort on the design of an adaptive voter be restricted to that of monitoring the state of the art in device development and to begin detailed circuit design when more suitable devices become available.

2. Threshold and Dynamic Restorers

The majority voting class of threshold restorers are the most commonly used restorers in present technology. Because the majority voter requires a majority of correct inputs to provide a correct output, its error-correcting capability is limited. Since many circuit failures result in steady-state outputs, restorers which detect only changes in input states offer the capability of deriving a correct output with less than a majority of working inputs. Restorers which detect changes in input states are referred to as dynamic restoring circuits.

The mission of this part of the Failure Free Systems Study has been to evaluate the potential usefulness of one proposed dynamic restoring circuit implementation, the Transor.

The results of section IV have shown that there are certain environments in which Transor can be used to advantage in improving system reliability. For example, the maximum error restoring capability of Transor is shown to be $R-1$ failures of R redundant lines in an environment free from transitional failures. This is a significant improvement over the majority threshold restoring capability under the same conditions. There is need for caution, however, for in environments where symmetrical transitional errors are possible, error correlation may make Transor performance inferior to threshold.

During the course of the study of Transor Restoring Circuits, a new class of restoring circuits was conceived. This class, called "Hamming Distance Restoring Circuits" is similar to Transor in many ways. It was compared with Transor analytically and by simulation. From the results obtained by manipulating the analytical reliability expressions for the Transor and Hamming Distance Restoring Circuits, it may be concluded that the output of a Hamming Distance Circuit is more reliable than that of the Transor in order-five redundant systems. This conclusion holds for any ratio of steady-state to transient error probability or any asymmetry (tendency toward "ones" or "zeros") of error probabilities.

From comparison of the simulation curves, it may be concluded that the threshold circuit is more reliable than either of the dynamic restoring circuits until the ratio of the probability of steady-state errors to the probability of transient error exceeds approximately seven to one. Above this ratio, the dynamic restoring circuit outputs are more reliable. Further comparison reveals that the difference in the reliability curves tends to stabilize or slightly decrease as the ratio becomes much larger than 7:1. The stabilizing effect is more pronounced as the order of redundancy is increased from five to seven.

Also, it may be concluded that in the early life, high reliability region with approximately a seven to one probability ratio, an order five system using Hamming Distance Restorers may be as reliable as an order seven system using threshold voters.

Since the improvement available from Transor is limited, and since the Hamming Distance Restorer is normally superior, further study of the Transor is not justified.

TASK 3 - SELF REPAIR TECHNIQUES

Before self-repairing systems can be implemented, many feasible switching strategies must be considered in an effort to determine the most effective manner to manipulate the redundant or "spare" blocks. The extreme complexity of the reliability expressions associated with these strategies has resulted in the use of a computer simulation program for comparing the effectiveness of the strategies. The present program includes subroutines for three classes of switching strategies. Each class subroutine contains a great deal of flexibility,

thereby including many individual strategies. This method facilitates easy comparison between members of a class. This comparison allows immediate elimination of many possible strategies which are obviously uneconomical. For example, the flattening out of the Percent of System Failed versus Spare Mobility curves indicates that none of the strategies on the flat part of the curves can be optimum strategies.

From the results of the simulation program, curves for Percent of Systems Failed versus Spare Mobility have been plotted for the Gamma Class Strategies. These curves have been referenced to that of a multiple-line majority voted system because this particular technique has been the most effective of the passive, failure masking, circuit level redundancy techniques. In all cases these curves show not only that great gains can be realized over the multiple-line redundant configuration, but that by far the greatest part of these gains are realized for the first few moves allowed to the spare function blocks. Beyond the range of relatively limited mobility, little or no gain in the average number of failures absorbed is realized by the additional mobility allowed to the spares. This is an encouraging result since the great majority of the gain due to self-repair can be retained without the use of an exorbitant amount of switching circuitry.

All of the computer simulation results have been based on the assumption that the switching circuitry was perfectly reliable. There is a need to determine the range of allowable failure rates which can be associated with each strategy for it to be of maximum effectiveness. These ranges should be studied as a function of the failure rates of the associated signal processor blocks. As a result, information specifying the optimum switching strategy corresponding to a given signal processor failure rate should be available before actual system designs are begun.

It has become obvious that many of the spare function blocks do not experience as many switching operations as they are capable of performing. When all spares are assigned mobility, those which use their mobility extend the life of the system substantially. However, in many cases when system failure has occurred, there are many spares remaining which have not been used to any great extent. In order to try to capitalize on this phenomenon, a class of strategies should be investigated which would assign different mobilities to the spare in a stage.

The curves show a very definite gain in reliability for the self-repair strategies over multiple-line redundant systems. The curves for the Beta Class strategies show an increase in reliability for each increase in "repair" capability. Strategy Beta-3 yields the highest reliability but even strategy Beta-1 shows a significant gain over the multiple-line system. The reliability curves for the Gamma Class show essentially the same result with respect to

the multiple-line case. However, investigation of the curves show that increasing the "repair" capability produces gains for the first few increases, after which the magnitude of the gain diminishes. These curves tend to bear out the conclusions drawn from Percent System Failed versus Spares Mobility curves which flattened out after a certain mobility was reached. The gains illustrated here must be considered as ideal because the switching circuitry for self-repair is here assumed to be perfectly reliable. More realistically, the gains obtainable will be a function of the switching circuit complexity and will not be as great as shown here.

Although little has been said about the physical switching techniques to be employed, it has been tacitly assumed that the failure detection and replacement circuitry would be combined as much as possible. It has been suggested that these two phases of the repair function might profitably be separated and made almost completely independent from a circuit viewpoint. This is another area which should be given careful attention.

None of the strategies considered so far have permitted spares to return to previous locations. It is possible that removal of this restriction might add to the failure absorption capability of a system. This area certainly should be explored further.

The Alpha class strategies have not been thoroughly investigated to determine the optimum degree of spare overlap (i. e. , two sets of spares serving some of the same functional region). The information from this investigation should influence the design of new strategy classes as well as indicating the optimum strategy for the Alpha class.

In general, investigations to date have shown that self-repair techniques can be much more powerful than presently available redundancy techniques. Further studies are expected to show effective ways to apply the techniques to real equipment needs.

Appendix 1

DESIGN AND TESTING OF REDUNDANT SYSTEMS

by

H. Brinker

A. R. Helland

September 1963

ABSTRACT

This report describes the results of the study on the implementation of majority logic redundancy. Most of the work concerns spaceborne systems, but some portions are more applicable to ground support equipment. The report is concerned with the initial design of the system as well as the testing of redundant systems.

The possible use of magnetic logic to reduce the total power consumption and provide non-volatile storage is discussed. Magnetics seems to be most useful for non-volatile memory and simple forms of logic where the data rate is very low. Various types of semiconductor logic are described and compared for use in redundant systems. Integrated Diode-Transistor Logic elements are chosen as the most suitable for general use.

Several methods of testing redundant systems are discussed and described in the section on detection and location of failures. Various solutions to the failure detection problem are discussed in this section. Some are more suitable for simple failure detection; others also provide information concerning the location of any failures. It is shown that maintenance of a redundant system greatly increases system reliability and reduces the test equipment and operator skill which are usually required to maintain a conventional system. Techniques are described which permit a major portion of the maintenance to be performed during normal system operation.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. MAGNETIC LOGIC	5
A. Introduction	5
B. Dynamic Storage and Sequential Logic	6
C. Hybrid Devices	7
D. All-Magnetic Logic	13
E. Summary and Conclusions	21
III. SEMICONDUCTOR LOGIC	25
A. Introduction	25
B. Classification of Basic Types of Logic	26
C. Comparison of Logic Types	31
D. Description of Logic Types	34
E. Logic Selection	41
F. Majority Voter Design	43
IV. FAILURE TESTING OF REDUNDANT SYSTEMS	45
A. Introduction	45
B. Singular Rank Testing	61
C. Interwoven Rank Testing	71
D. Circuit Implementations	79
V. SUMMARY & CONCLUSIONS	84

LIST OF FIGURES

Figure	TITLE	Page
1	OR Gate	9
2	Negation	10
3	Block Diagram, AND Function	11
4	SRI MAD Shift Register	14
5	Device 2 Flux States	17
6	Device 2 Shift Register	19
7	R-TL Resistor-Transistor Logic (+NOR)	27
8	DC-TL Direct Coupled-Transistor Logic (+NOR)	28
9	R-DC-TL Resistor-Direct Coupled-Transistor Logic	28
10	NS-DC-TL Non-Saturated-Direct Coupled-Transistor Logic	29
11	D-TL Diode-Transistor Logic (+NAND)	30
12	NS-D-TL Non-Saturated-Diode-Transistor Logic	30
13	T-TL Transistor-Transistor Logic	31
14	Speed-Power Performance	37
15	Majority Element with Input Isolation	43
16	Reliability of Conventional vs. Redundant Systems	45
17	Singular Rank Testing	62
18	Interwoven Rank Testing	73
19	Interwoven Rank Testing	74
20	Signal Processor Output Control	80
21	Difference Detector	82

I. Introduction

Past studies of redundancy techniques and consideration of the basic characteristics of some redundancy techniques have yielded interesting insights and problems. Many of these considerations are in the area of engineering method. Others concern the design of redundant systems with high reliability and other desirable characteristics. This section is intended to review some of these considerations and to preview some of the thoughts behind the discussion in later sections.

The report itself deals primarily with some of the problems which are encountered in designing and testing useful redundant digital systems. Some of these problems are at least comparable to non-redundant design; others are rather unique to redundant systems. Possible solutions for these problems, as well as more detailed problem descriptions, are contained in appropriate sections of the report.

Circuit and system design must reflect the fact that redundancy is only a tool to realize reliability. The proper use of redundancy is often a more efficient and powerful technique to realize a reliability requirement than are the more conventional techniques such as conservative design or component selection. Redundancy is, however, most powerful when used in conjunction with techniques that increase basic reliability.

It is important to recognize that a redundant system is expected to operate with relatively large numbers of random failures. Since conventional systems usually fail when any of their parts fail, it is relatively unimportant what effects these failures have, except when repair is desired.

Circuits for redundant systems, however, must be designed so that the effects of individual component failures are minimized, and usually limited to the circuits in which the failure occurs. This does not imply, however, that redundancy includes "useless" parts. Each part of the system must contribute to the assurance that the system will perform all of its functions properly.

The use of redundancy will alter the characteristics and performance of the system. Redundancy will usually increase design complexity, power requirements and dissipation, signal propagation time, size and weight, number of interconnections, and initial cost. Redundancy, therefore, emphasizes the need for continuing development of low-power circuitry, micro-miniaturization, and interconnection techniques. The type of circuitry which is used to implement a redundant system must be carefully chosen to meet the system requirements without incurring excessive costs. Whenever there is a need for high reliability, the circuitry should be chosen to have a high basic reliability, low sensitivity to parameter variations, and low power dissipation to minimize temperature stress. In addition, specific systems have special requirements which must be considered in the system design as well as the choice and design of the circuitry. For example, the total available power is often severely limited for spaceborne equipment, although the processing rate is usually quite low. It is usually desirable to provide some means of testing to verify that all parts of the redundant system are working to insure that all of the reliability initially designed

into the system is available for the duration of the mission. The system and the circuitry therefore must be designed so that accurate and meaningful tests may be applied to verify that the parts are working. When extended lifetime is desired and repair is possible, a redundant system may be systematically repaired to greatly increase the expected time between system failures. If a system is completely repaired prior to each mission in which it is used, it will exhibit the high mission reliability characteristic for each mission. Such systems must be designed so that complete, efficient tests may be periodically applied to these systems which will verify that all the parts are working properly, or that will facilitate maintenance procedures which will return the system to the initially perfect condition. It is important for this type of maintenance that all failures be detectable, otherwise these undetectable failures will tend to accumulate. These accumulated failures will eventually tend to dominate the system behavior by causing additional system failures.

Many failures may be detected as they occur in a redundant system. These may be repaired while the system is in operation to obtain a very low system failure rate compared to the failure rate for the parts of the system. Periodic maintenance must be performed in addition to the continuous monitor and repair described above to detect those failures which cannot be detected during regular operation of the system.

Systems which will be maintained must therefore be designed both with the capability for detecting all failures and facilitating the maintenance and repair procedures. With proper design, many of these failure

detection, maintenance and repair procedures may be accomplished during operation of the system.

The following sections of this report will discuss the problems associated with circuit design, choice of the type of circuitry, failure detection, and maintenance of redundant systems. This report describes the results of the study of these problems and possible solutions. The results are summarized in the Summary and Conclusions section of this report.

II. Magnetic Logic

A. Introduction

The past decade has witnessed the development of a variety of magnetic devices suitable for performing storage and logic in digital computers. Perhaps the most important application of magnetics to digital technology has been provided by the development of large capacity, random access memory systems composed of ferrite cores. Advances in techniques for performing logic have received some attention, but to date magnetic logic does not appear to be widely accepted as a superior replacement for the conventional transistorized counterpart. This general reluctance to utilize the special attributes of magnetic logic is often justified by several difficulties inherent to the device characteristics and system configuration.

Much of the magnetic logic research has been motivated by the potential ability of magnetic devices to provide higher reliability at lower cost while consuming negligible standby power. These attributes are understandably important in any large electronic system, especially in space applications where reliability must be high and available power is invariably low. To evaluate the potential ability of magnetic logic schemes to provide these advantages a discussion of some of the more promising approaches appears to be in order. An all inclusive survey and treatment of the myriad of suggested approaches could easily fill a book.* It appeared

* Edited by Meyerhoff, A. J., Digital Applications of Magnetic Devices, New York; John Wiley and Sons, Inc., (1960).

reasonable therefore to restrict the detailed discussion to the more popular approaches and to provide references for other. Of particular interest are those devices which utilize magnetic components which are either commercially available or in an advanced state of development.

B. Dynamic Storage and Sequential Logic

The state of a magnetic device is determined by the direction of remanent flux. Information stored is not directly accessible and a clock or read pulse must be used to determine the state. The read process in most schemes also destroys the information which was stored. An output signal is available only for that portion of the read cycle during which dynamic flux change is in progress and thus level output and asynchronous operation is not obtainable. The ripple-carry binary counter, the parallel adder, and many familiar digital configurations are not directly amenable to magnetic implementation. In contrast, the powerful combinational logic approach utilized in conventional computers consists of a cascade of compatible logic modules which form complex functions simultaneously during the interim between clock pulses. In a magnetic logic machine using dynamic logic this is not possible and operations involving OR, AND, transfer, buffering, negation and delay require several clock periods to generate a particular function. This step by step process usually consumes considerable time which may be further extended if the magnetic logic modules are limited in fan-in and fan-out and thus require additional operations.



C. Hybrid Devices

The principle involved in using square loop material to store a remanent flux has been known for some time. With the development of small torroidal structures employing sintered ceramic ferrites and ferromagnetic tape materials, magnetic devices began to demonstrate practical utility. The magnetic shift register has received the most attention primarily because of its general utility and simple configuration and has been the subject of much of the magnetic literature. Although playing an important part in most digital systems, several additional devices are required in order to provide the variety of logical operations required by typical computer systems.

The task of performing general logic requires circuitry capable of being arranged to perform any Boolean output function of a set of input variables. In order to provide this operation a complex function is usually formed by using logic modules to perform OR, AND, negation, storage, delay, etc. If gates are to be connected in various configurations the devices used must provide a clearly identifiable "1" and "0" state, unilateral information transfer and the capability for fan-in and fan-out. To meet these requirements with magnetic devices has not been an easy task.

A major difficulty which impeded rapid development of devices to meet these requirements has been the inherent bilateral nature of simple magnetic structures. In the early devices this was largely overcome by combining diodes with simple torroids to achieve unilateral information

flow. Obvious limitations in impedance levels, fan-in and fan-out drive capabilities necessitated in many cases the further inclusion of resistors for tailoring impedance levels, capacitors for temporary storage and transistors for power gain. Although this hybrid logic approach led to the development of a number of clever magnetic devices, the potential of achieving high reliability at low cost is seriously challenged by the requirement for using non-magnetic components and the more complex wiring and system organization which becomes necessary. An excellent survey of a wide variety of hybrid devices has been provided by Haynes.¹ One such approach, parallel transfer core-diode logic, will be used as a vehicle for describing the principles of dynamic logic and to indicate the operation of a typical practical device.

Shown in figure 1 is the OR gate, the simplest of logical functions which may be implemented with magnetic cores and diodes. The  and  notations denote cores of the same rank, i.e. threaded by a series connected, current driven clock line. The two phase clock system effects readout and transfer of data by driving the core to the "0" state. If a core was previously in the "1" state the clock, in driving the core to the "0" state, causes the core to switch and provides an output sufficient to drive the next core to the "1" state. If a core was previously in the "0" state a negligibly small output occurs when the clock drive is applied. Diodes are shown to prevent output loading when a core is being set. Additional components such as resistors for tailoring impedance levels and

diodes to prevent reverse data transfer may be required in a practical design. It should be noted also that the core output windings must contain more turns than core inputs in order to allow a transmitting core to set a receiving core, which also tends to prevent reverse data transfer.

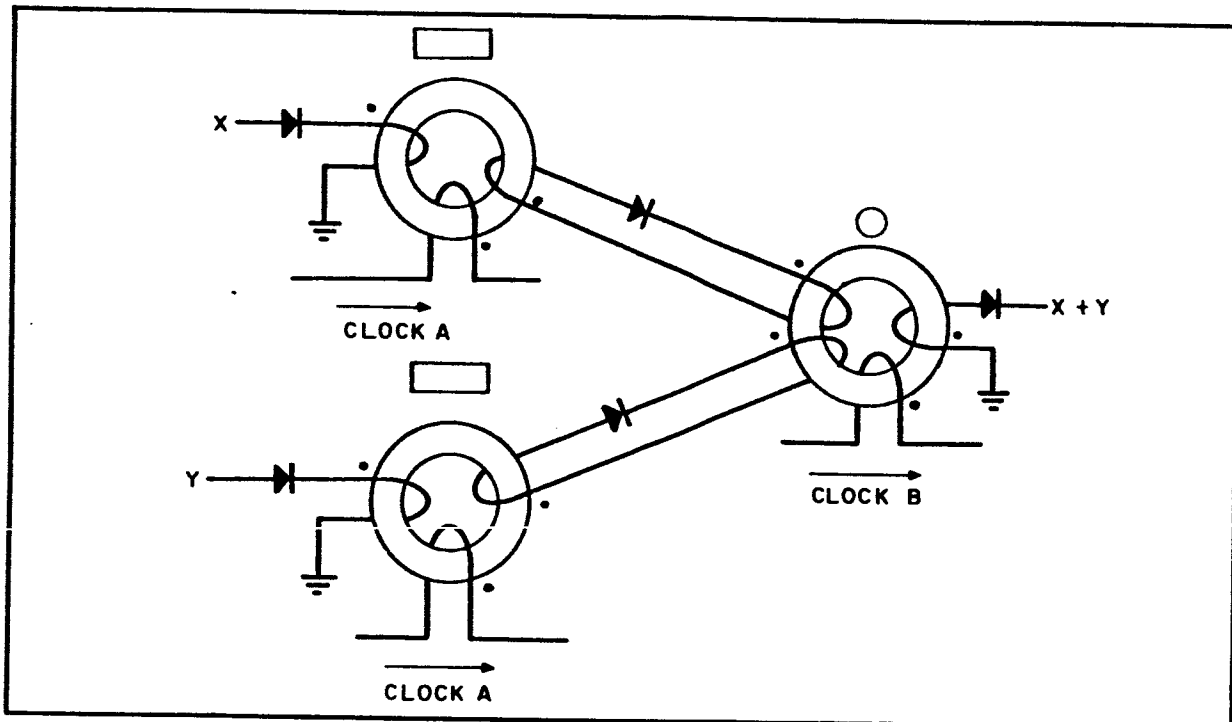



Figure 1 OR Gate

Operation is initiated by reading inputs X and Y into the  cores. The phase A clock then transmits the state of each of the input cores into a dual winding storage core. If the storage core was set by

any of the transmitting input cores, a readout signal is generated when the storage core is reset by the phase B clock.

The AND function is not as easily implemented unless a coincident current threshold technique is employed to set the storage core. This technique does not appear to be sufficiently reliable however, due to the associated threshold and drive tolerances normally encountered in a typical system. A more conventional system employs the principle of logical negation in combination with the OR gate to provide the AND function. For example, consider the negation arrangement of figure 2.

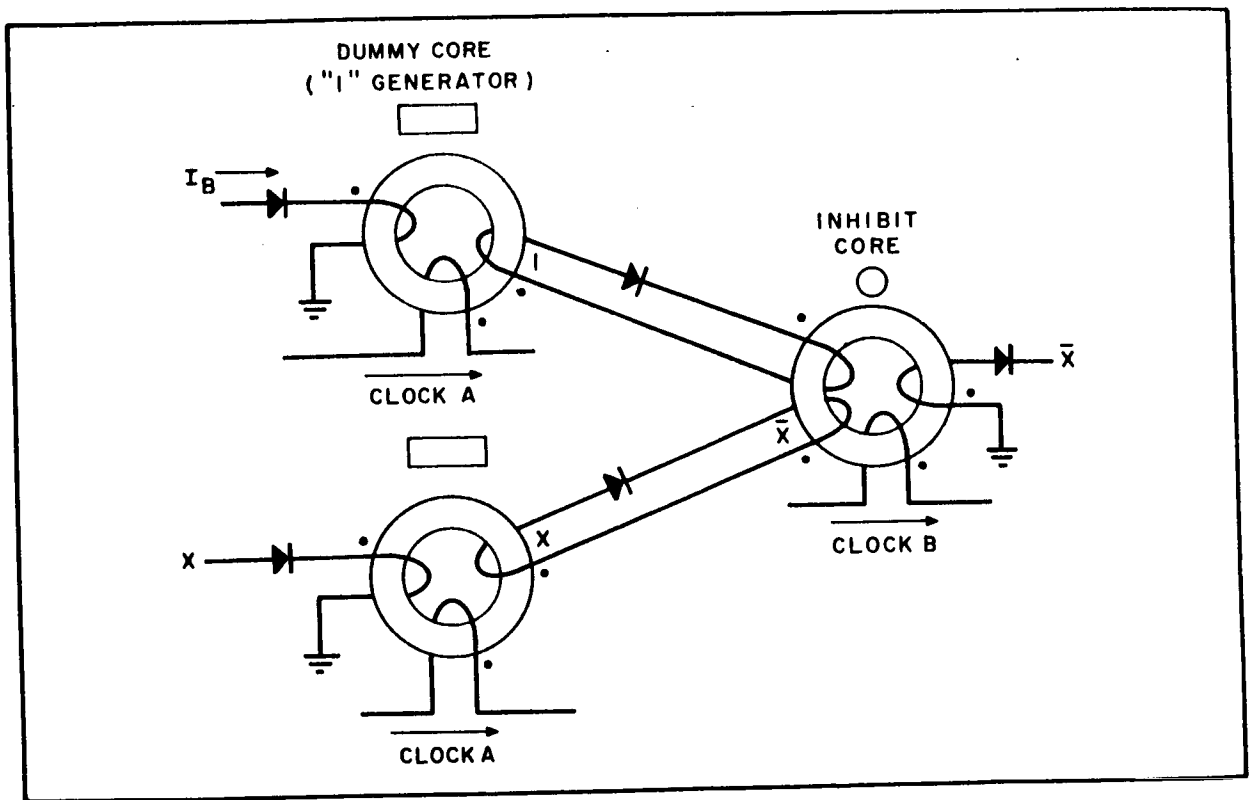


Figure 2 Negation

The upper core is used as a "1" generator which in the absence of an input from the X core causes the inhibit core to be set by the phase A clock. The phase B clock will then generate an output whenever the X signal is absent and thus represents the negation of the input. When both the "1" generator and X input signal appear simultaneously at the inhibit windings they effectively cancel each other and the inhibit core remains in the "0" state. The phase B clock in driving the inhibit core to the "0" state will not generate an output signal for this case.

The principle by which the AND function may be performed is based on the well known logic relation $\overline{\overline{X} + \overline{Y}} = XY$. A block diagram of a typical AND gate scheme is shown in figure 3.

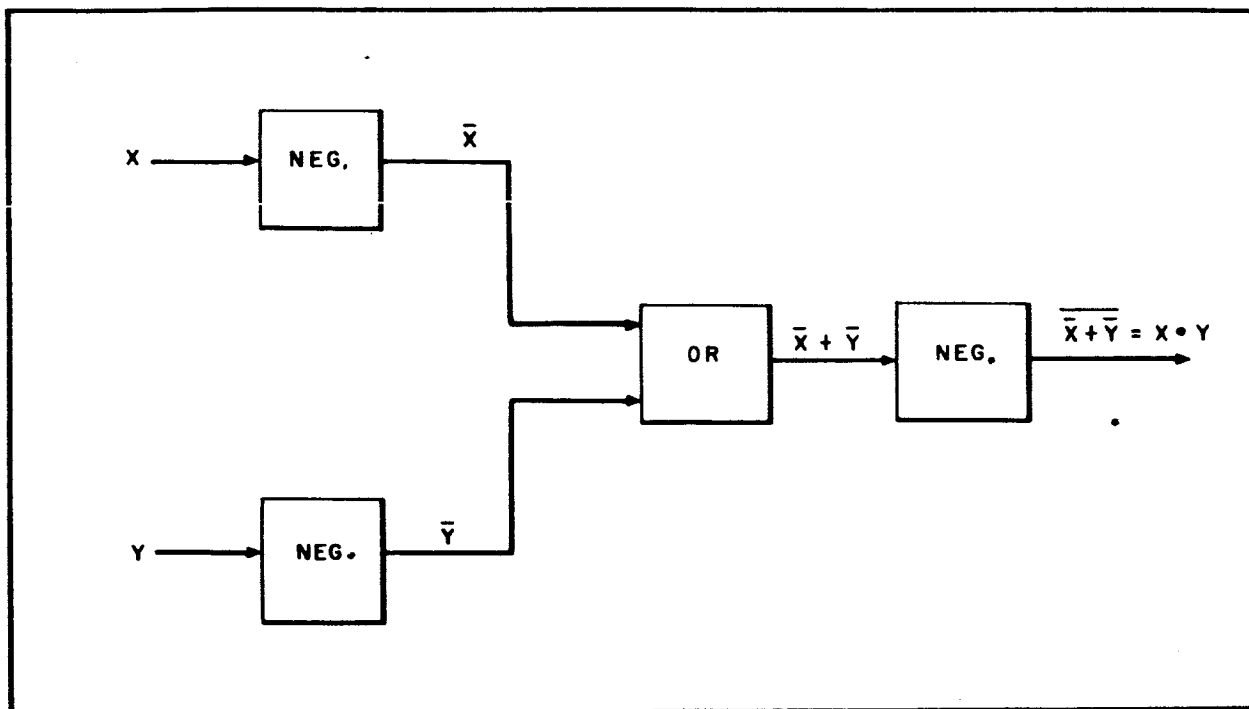


Figure 3 Block Diagram, AND Function

Since each of the logic modules require two clock periods and each operation is performed in sequence, the output signal is seen to appear six clock periods after the inputs were applied. If the resultant output of the AND function is to be further combined with other AND-OR operations it becomes evident that the total number of clock periods required may become prohibitive.

In view of the system complexity and speed limitations suggested by the simple example described, magnetic logic is seen to introduce problems of system organization which are alien to conventional DC level logic. As far as cost and reliability are concerned, the prospect of winding cores with several turns and the large number of cores and connections required do not appear to provide a significant cost advantage. In the hybrid approach the use of additional components such as diodes and resistors appear to seriously negate the basic reliability inherent to the magnetic material. These difficulties notwithstanding, several companies are active in the manufacture of magnetic logic modules. The major emphasis has been placed on the usefulness of the magnetic shift register to provide cost, size and power advantages over the conventional approach. Magnetic shift registers employing the hybrid approach have been successfully applied to a wide range of airborne equipment. Sequential programmers, counters and timers operating at low clock rates represent the majority of applications. When operating at shifting rates higher than 10 kc however, the

advantage that the magnetic shift register has in consuming negligible standby power is obscured by a power requirement which is often greater than the solid state counterpart. A leading supplier of hybrid magnetic logic modules and shift registers is currently marketing a 10 bit shift register which requires a maximum average power of .4 watts to operate at 10 kc and 3.7 watts at 750 kc. Since it appears reasonable to assume that these power requirements are reflected also to general logic systems, the application of hybrid magnetic logic to power-limited environments is limited to systems whose shift rate is very low.

D. All-Magnetic Logic

The obvious limitations of the hybrid approaches in reliability and cost has to some extent motivated an effort to develop systems using only magnetic material and connecting wire. Several novel approaches were developed which made use of magnetic device geometry to achieve coupling isolation, flux gain and unilateral information flow. One of these devices is the Multi-Aperture Device (MAD),^{2,3} a three aperture ferrite structure similar to the Transfluxor.⁴ Input-output isolation is possible because the flux stored around the minor output aperture may be sensed non-destructively without affecting stored flux about the input aperture.

Shown in figure 4 is a typical MAD shift register developed at Stanford Research Institute.

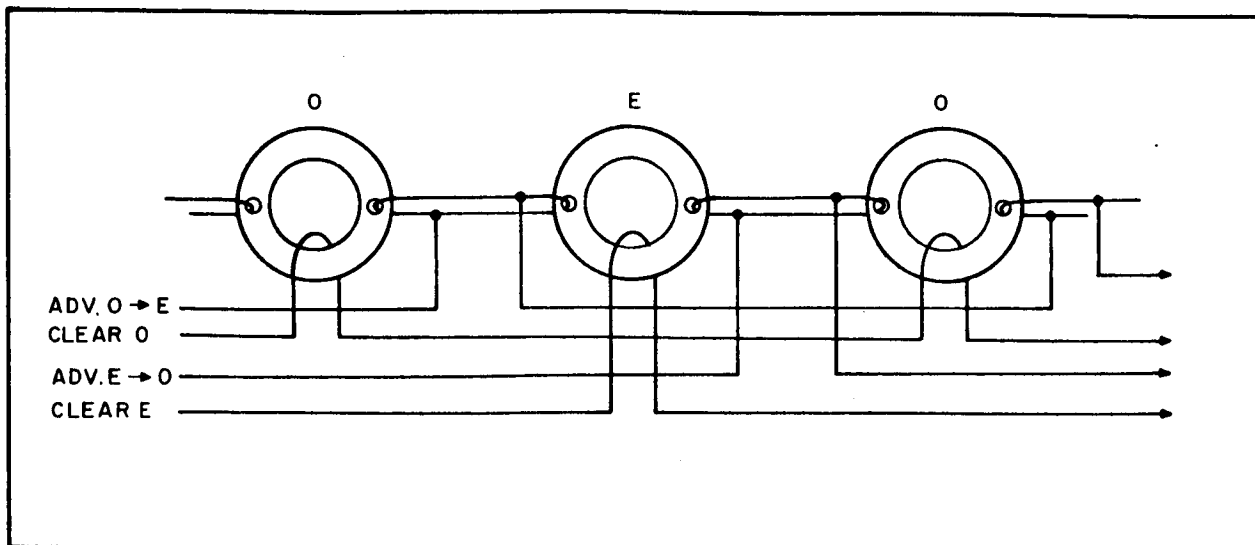


Figure 4 S.R.I. MAD Shift Register

An advance current is applied to the parallel connection of output and input aperture windings in order to effect information transfer from the transmitting core to the receiving core. In accordance with the state of the flux stored around the transmitting aperture and the resultant magnetic threshold thereby established, the advance current will divide between the input and output windings. If the transmitting aperture is in the "0" or cleared state the advance current will divide equally thus not exceeding the magnetic threshold of either apertures. If a "1" were stored the output aperture with its lower threshold is swamped by the advance current and the transmitter switches flux locally about its output aperture with low values

of current. By voltage or impedance steering the majority of advance current will flow through the receiver input aperture causing it to exceed its setting threshold and be set. In time as the flux switching is completed, both currents will return to their nominally equal values.

Since the read-out and transfer process is nondestructive to the state of the core, a clear line threading the major aperture is required to return the core to the reset condition. In order to provide information flow from left to right a basic four clock cycle is required with the following sequence:, ADV.O \rightarrow E, CL.O, ADV.E \rightarrow O, CL.E, ... The ADV O \rightarrow E pulse switches flux locally about the output aperture of the O element and causes the E element to be set. The CL O pulse then clears the O element and in so doing switches flux through the output winding. This results in a loop current flow that negatively sets the E element receiver without affecting the flux state about the output aperture of the E element. Note that neither the ADV. O \rightarrow E nor CL. O pulse causes any flux to be switched in the output leg of the E element thus eliminating the need for a diode to prevent backward data transfer. In this manner unilateral data transfer is possible using only MAD devices and conducting wire.

Thus far the discussion has been devoted to techniques for achieving unilateral data transfer with the S.R.I.-MAD approach. The problem of achieving reasonable flux gain and fan-out is one which could not be solved

in a practical sense with the simple transfer scheme previously discussed. H.D. Crane has done much of the work in arousing interest in the all-magnetic MAD approach. In a paper⁵ describing the design of a moderate sized computing system using S.R.I.-MAD devices however, the basic transfer gate had to be seriously modified in order to operate in the system. Problems inherent to the flux threshold relationship between receiving and transmitting apertures, flux gain, fan-out as well as flux decay and build-up in circulating loops made such modifications necessary. As a consequence the revised gate module required flux doubling and clipping operations in addition to the previously described clear and advance cycles. The complexity involved in the resultant device implementation appears to be a serious encumbrance. The system chosen to demonstrate the ability of all-magnetic devices took the form of a decimal arithmetic unit with the ability of performing addition, subtraction, and multiplication. The system was made exclusively of modules which perform either the two input OR function or the two input OR with negation (NOR).

Rather than describe the complex details of the S.R.I.-MAD logic gates it appears more reasonable to present an alternate approach to the design of MAD devices developed by Company I. In this approach a priming operation is performed to reverse the flux stored about the transmitting aperture prior to readout. The readout process in this case is destructive and resets the core. The priming operation provides an adequate flux level which, when reversed by the clear or transfer

operation, delivers an output pulse to set the next core through its major aperture. Since data flow is from minor aperture to major aperture and since the state of a core is not disturbed by reverse currents flowing through a minor aperture, the possibility of reverse data flow is prevented.

The flux conditions present for the various states of a typical MAD element of this type (referred to as Device 2) is shown in figure 5.

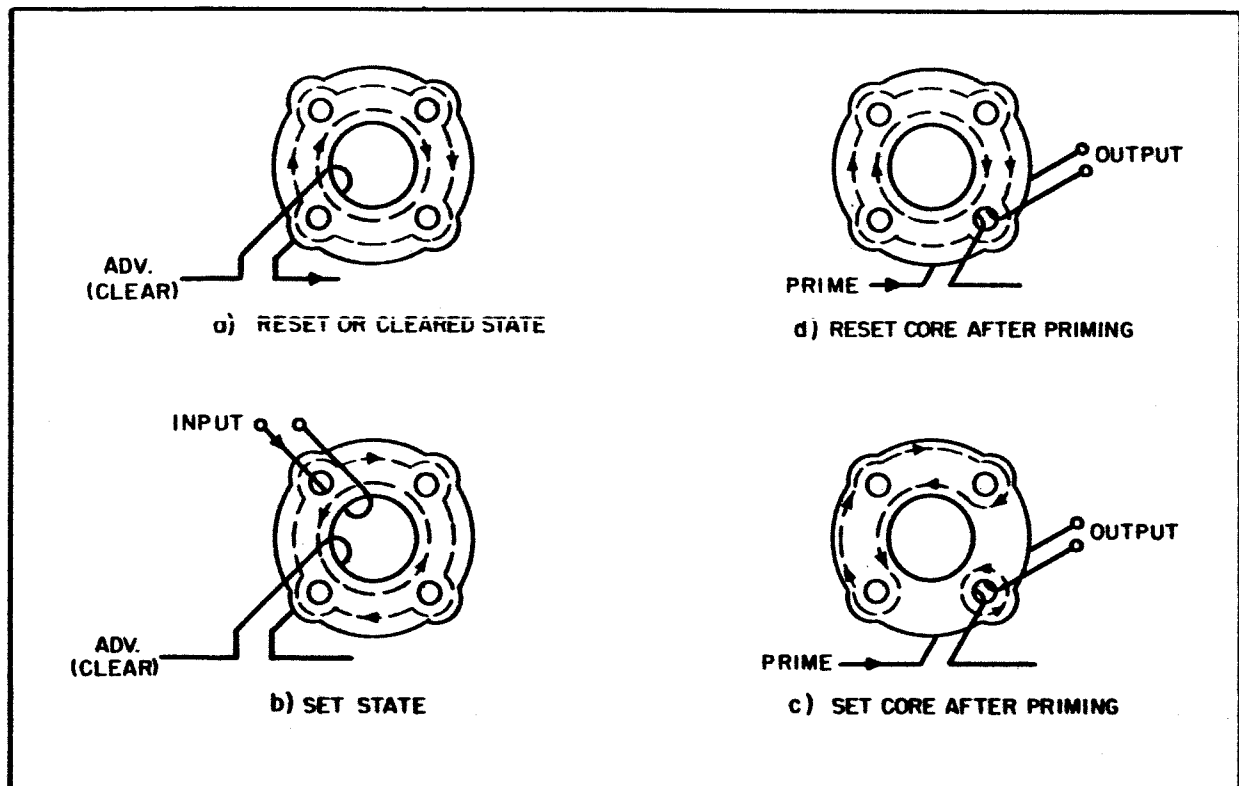


Figure 5 Device 2 Flux States

In the cleared state (figure 5a) the core is saturated in the clockwise direction by a previously generated advance current which threads the major aperture. Upon application of an input signal threading the inner portion of the major aperture, the flux nearest the major aperture is reversed thus providing the set condition shown in figure 5b. This read-in operation does not affect the flux linking the output aperture and thus a diode is not required to block data transfer to receiving cores. In order to obtain an output from a properly set core it is necessary to provide a prime current as shown in figure 5c to reverse the flux stored about the output aperture. Priming current is of a lower magnitude than the advance current and because of its slow rate of change is not sufficient to cause the core linked by the output winding to be disturbed. Once a core has been set and primed, the application of an advance current causes a flux reversal about the output aperture. This in turn, provides an induced voltage of sufficient magnitude to drive the next core to the set condition. If the core was initially in the reset condition it will remain in this condition after priming (figure 5d). For this case, the application of the advance current does not provide a flux reversal and thus no output occurs.

Device 2 elements may be connected in a variety of shift register configurations including parallel input-parallel output, parallel input-serial output, serial input-serial output, etc. Such shift registers take the form of 2 core-per-bit arrays and require a two clock system in combination with

a priming source. A typical serial input-serial output shift register section is shown in figure 6.

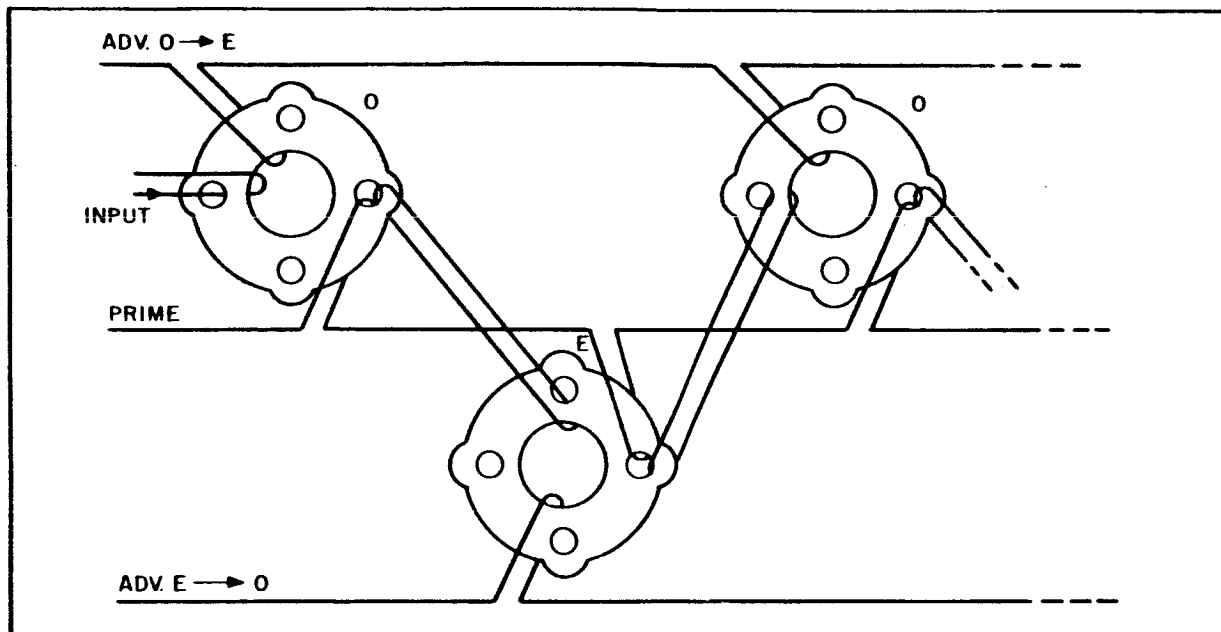


Figure 6. Device 2 Shift Register

The propagation of a "1" from left to right proceeds by activating clock and prime signals in the following sequence: ... PRIME, ADV O→E, PRIME, ADV E→O, PRIME, ADV O→E, AMP-MAD shift registers require relatively high values of pulse current for performing advance, prime and set operations. Nominal operating level for the advance current is 2 to 3 amperes in a typical design. Prime and set pulse currents are lower being 100 ma and 250 ma respectively. Because of the requirement for slow priming and in order to keep average power dissipation at reasonable levels, these

shift registers are limited to repetition rates of 10 Kc. A typical driver, which utilizes a capacitive storage-discharge scheme and dual Shockley diodes for triggering the advance currents, requires an average power of 5.3 watts to drive a 10 bit shift register at 10 Kc. A 10 bit shift register with its associated driver requires a package occupying approximately 9 cubic inches.

The implementation of general logic operations using MAD devices is not easily accomplished, due to the difficulty of achieving logical inversion and reasonable fan-out without an imposing complexity. The treatment of much of the general logic capabilities of MAD devices is reported in rather implicit terms by the current literature. The OR function may be provided relatively simply by threading additional windings about the input aperture if care is taken in preventing reverse information transfer. The negation operation may be achieved by extending the current inhibiting and "one" generator technique described in the hybrid approach to the MAD topology. Perhaps the most difficult problem which faces the all-magnetic logic designer is that of providing fan-out. This arises from the fact that all the power which is used to provide inputs to receiving cores comes from the clock source. Power gain in the ordinary sense is not available except in those hybrid schemes which use transistors to provide regeneration. A MAD device with a reliable fan-out of two is sufficient, however, to allow the performance of general logical operations requiring much greater fan-out. This may be accomplished by utilizing additional clock pulses to

sequentially transfer data in a "tree" wiring arrangement until the original single core data is available simultaneously in several cores. As far as fan-out is concerned, it appears that the hybrid approach using transistors provides an important advantage over the all-magnetic techniques which necessarily require considerable device and system complexity to achieve the same result.

E. Summary and Conclusions

The foregoing description of magnetic logic has not attempted to describe the variety of possible approaches. The techniques for accomplishing general logical operations have been implicit, reflecting the treatment of the current literature. Examples from two general classes of magnetic devices have been described to provide a basic understanding of the techniques involved. If the approaches described may be regarded as typical, then some conclusions about their utility may reasonably be expected to apply in a general sense.

Information regarding transfer and shifting operations are covered in considerable detail by current literature, but the treatment of general magnetic logic schemes has been seriously neglected. This suggests the degree of difficulty which has been encountered in the design of practical devices. Complex clock programming and device configurations are necessary to achieve operations which conventional designers have come to consider as

trivial. In general, magnetic devices do not display a natural ability for performing logic. The primary attribute of magnetic devices is that of non-volatile storage, the ability of a core to remain in a particular state indefinitely without further application of energy. This feature is an important consideration in power limited environments such as space vehicles where the standby power between clock pulses may be made to approach negligible values. If the clock processing rate exceeds approximately 10 Kc however, the average power required often exceeds that of a conventional transistorized counterpart. This limits the application of magnetic shift registers, timers, etc. to equipment with low clock rates.

Recent advances in low power microminiaturized devices are seriously challenging the magnetic attribute of zero standby power while providing higher speed, smaller size and the greater utility of combinational DC logic. NASA's Lewis Research Center is sponsoring much of the work in this important area. Operating speeds of several newly developed circuits are approaching 100 Kc at power levels in the microwatt range. A complete logic system with a power consumption of 10 microwatts per stage is anticipated for space application using micropower logic circuits. With the basic reliability of microminiaturized devices constantly improving by virtue of an industry-wide effort, the role of magnetic logic appears to be fading.

Another advantage claimed for magnetic devices is the reliability inherent in the use of magnetic material and connecting wire. It is assumed here that magnetic parameters affected by temperature have been compensated

for by proper design and that clock current amplitude and rise time are within the limits of proper operation. Under these conditions the basic mechanism of magnetic storage and switching appears devoid of any known failure mode. This reliability is however obscured by the large number of connections required by the device configuration and the complexity inherent to the system organization. The reliability of a magnetic system depends upon the connective paths and the clock pulse drivers.

Simplicity and low cost is often claimed as a virtue for magnetic devices because of the simplicity and cost of the basic cores utilized. It should be noted however that the task of providing several turns about the various apertures and connecting cores in a configuration to perform the basic logical operations of AND, OR and negation is not generally amenable to automated assembly. The extensive amount of hand wiring and soldering appears to represent an item of considerable cost.

The physical size of magnetic devices are generally one or two orders of magnitude larger than their microminiaturized counterparts. Advances in thin film magnetic logic hold some promise for a significant size reduction, but developments in this area have not been extensively reported to date.

The flexibility of magnetic devices is seen to be severely limited by the dynamic logic approach and the difficulty of achieving reliable fan-out in the absence of active devices. The flexibility of conventional

DC logic systems is evidently superior because of the power gain and the inherent signal level standardization.

After considering the attributes of magnetic devices for performing general logic, the popular core techniques do not appear to provide an evident superiority in power consumption, reliability, simplicity, cost, size and flexibility over the conventional solid state circuit approach. Indeed, the requirements of performing the logical operations characteristic of digital computers appear to be at variance with the capabilities of magnetic logic. The applications which are best suited to magnetic implementation are those in which the operations to be performed are not clearly separated into "logic" and "memory". A strong case can be made for magnetic circuits applied to the performance of integrated storage and transfer operations required by a variety of digital processing functions. Most appropriate are the low speed operations inherent in input-output, interface and peripheral equipment. Typical applications include shift registers, programmers, timers, sequencers, etc. where the magnetic modules perform entire functions rather than discrete operations of storage and logic. In these special applications where speed is low, the advantages in simplicity, reliability, cost and power to be gained through the use of magnetic circuits should not be neglected. In general applications, however, the presently developed magnetic circuits do not appear satisfactory due to the several problems inherent in their use.

III. Semiconductor Logic

A. Introduction

In contrast with the numerous disadvantages and the general unavailability of magnetic logic devices, conventional semiconductor logic has been used widely. Logic modules are commercially available for construction of general logic systems. Integrated semiconductor circuits offer an order of magnitude reduction in size compared to magnetic logic modules; they do not require high voltage or high peak power pulses. They operate at frequencies many times greater than comparable magnetic logic requiring the same average power, and provide the convenience of steady voltage outputs.

Integrated semiconductor circuits offer a significant size and power reduction compared to discrete component semiconductor circuits. The rapid acceptance of integrated and semiconductor logic elements attests to the advantages of their use. Therefore, integrated circuits have been chosen as more suitable for spaceborne digital applications than the discrete component circuitry. The circuit design problem is then translated to the problem of the choice of suitable types of circuitry and logic. A variety of such elements is available with predictable characteristics for a wide range of operating environments. The selection by the Air Force of integrated circuitry for use in the improved Minuteman is a significant factor in the availability of reliable integrated circuits and appropriate reliability data. There is also a large amount of government

and industry effort devoted to research and development of new and improved : integrated circuits.

The low weight and power consumption of integrated circuits offers an important compensation for the increase in the number of circuits required for redundant design of spaceborne equipment. It is expected that advances in integrated circuit technology will allow more complex circuits to be included within a single package to further decrease size and weight. Integrated circuits also offer significantly improved reliability performance; it is expected that the reliability of single chip containing an entire function can be shown to approach that of a single discrete transistor. The low power consumption characteristic also tends to increase reliability by reducing temperature stress. The significant reduction in the number of interconnections is also an important factor in reliability improvement.

Most integrated logic modules are available in the form of a universal gate function (NAND or NOR). These logic elements are quite appropriate for the construction of the restoring function required for a multiple line majority voted redundant system. Several types of logic available for the universal gate function have been studied. Each basic type is described below; those commonly available are compared for suitability for use in spaceborne redundant systems. One of these is chosen as particularly suitable.

B. Classification of Basic Types of Logic

It appears that most of the common types of transistor logic (TL) may be classified according to three basic coupling schemes used for the

universal gate function. They are described below.

I. Linear impedance coupling to an input transistor may be used to form R-TL, as shown in figure 7. This type of logic is generally not available in integrated circuit form.

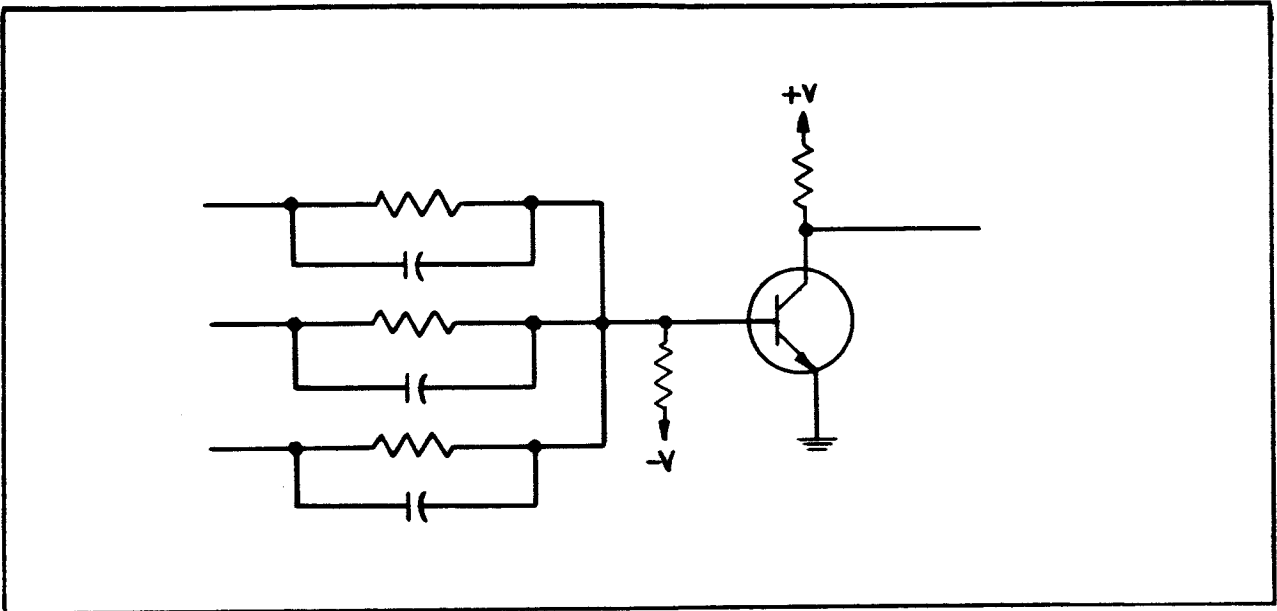


Figure 7 R-TL Resistor-Transistor Logic (+NOR)

II. Direct coupling to a multiple output transistor array (DC-TL), may be used as shown in figure 8. It is commonly used in the more practical modified forms, such as R-DC-TL (type II-A) shown in figure 9. An impedance is inserted in each input line to improve operational characteristics. Although this type of logic is sometimes referred to as resistor coupled-transistor logic, its operation is not the same as R-TL, described above.

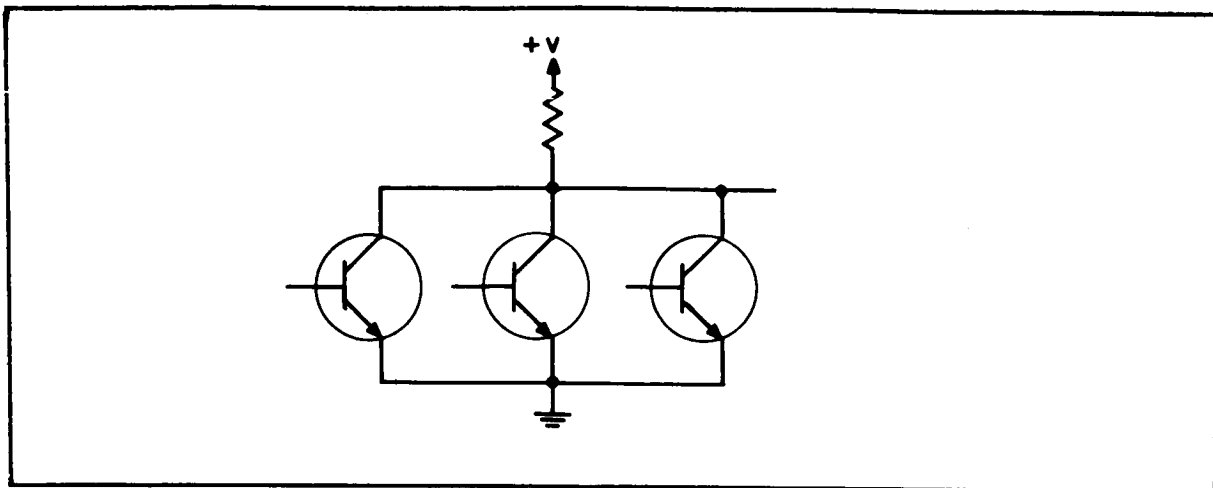


Figure 8 DC-TL Direct Coupled-Transistor Logic (+NOR)

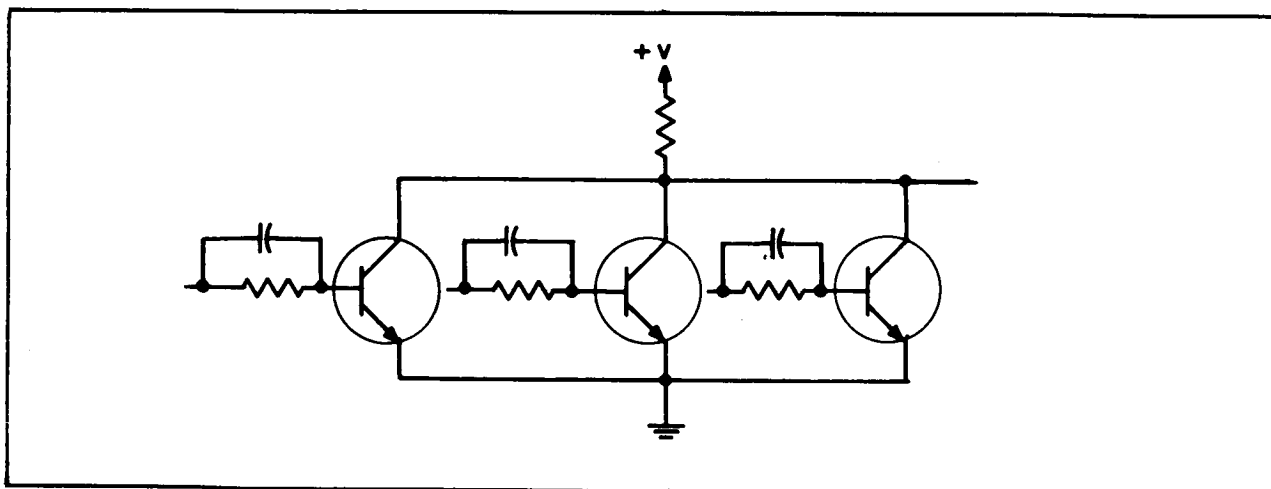


Figure 9 R-DC-TL Resistor-Direct Coupled-Transistor Logic

Type II-B coupling involves current switching and output buffering to prevent saturation of the input transistors. This type of logic is sometimes referred to as emitter coupled-transistor logic (EC-TL) or current mode-transistor logic (CM-TL). One type of non-saturated-direct coupled-transistor logic (NS-DC-TL), which uses an emitter-follower output buffer, is shown in figure 10.

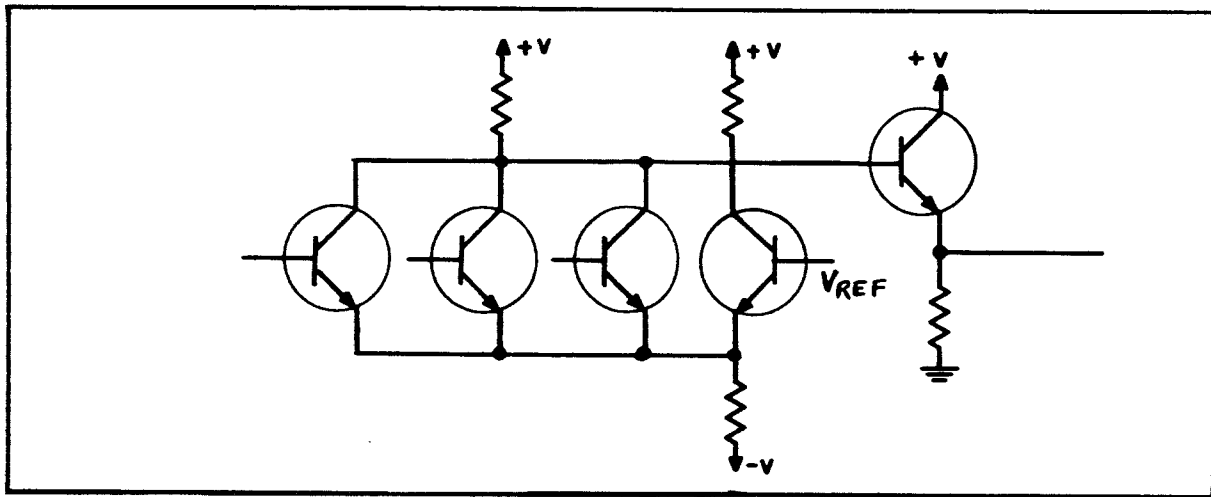


Figure 10 NS-DC-TL Non-Saturated-Direct Coupled-Transistor Logic

III. Diode coupling uses non-linear input summing to form the logical AND or OR function. The most common form of D-TL is shown in figure 11, which performs the positive logic NAND (AND-NOT) function. Saturation of the output transistor may be prevented by limiting the minimum saturation voltage, as shown in figure 12. This results in a more constant "zero" output voltage, and diverts excess base current to improve transient response.

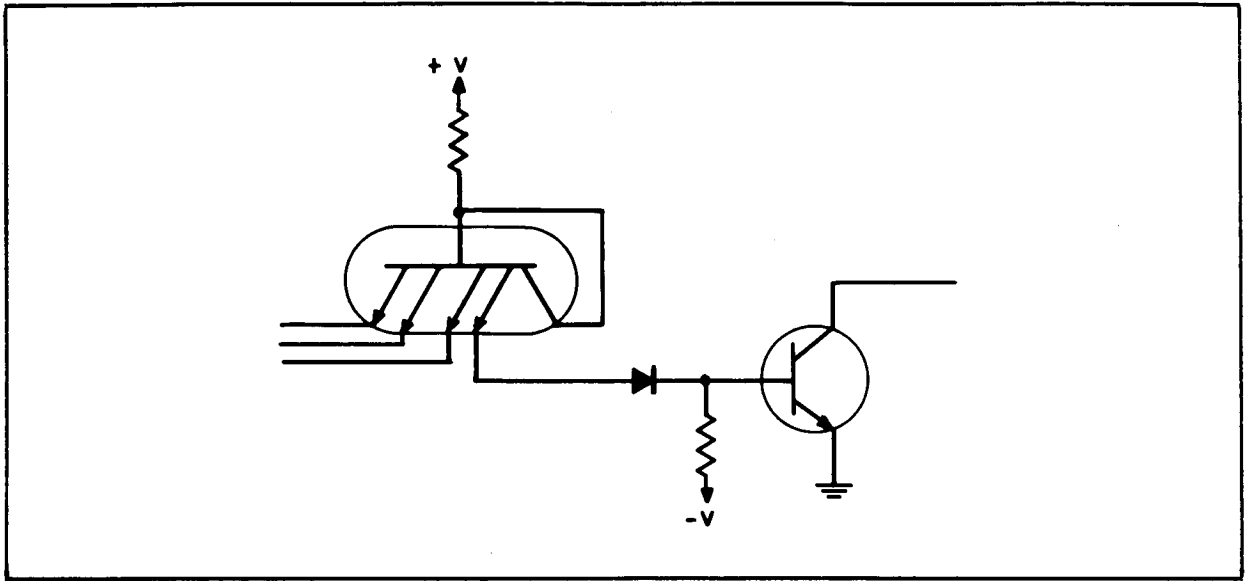


Figure 11 D-TL Diode-Transistor Logic (+ NAND)

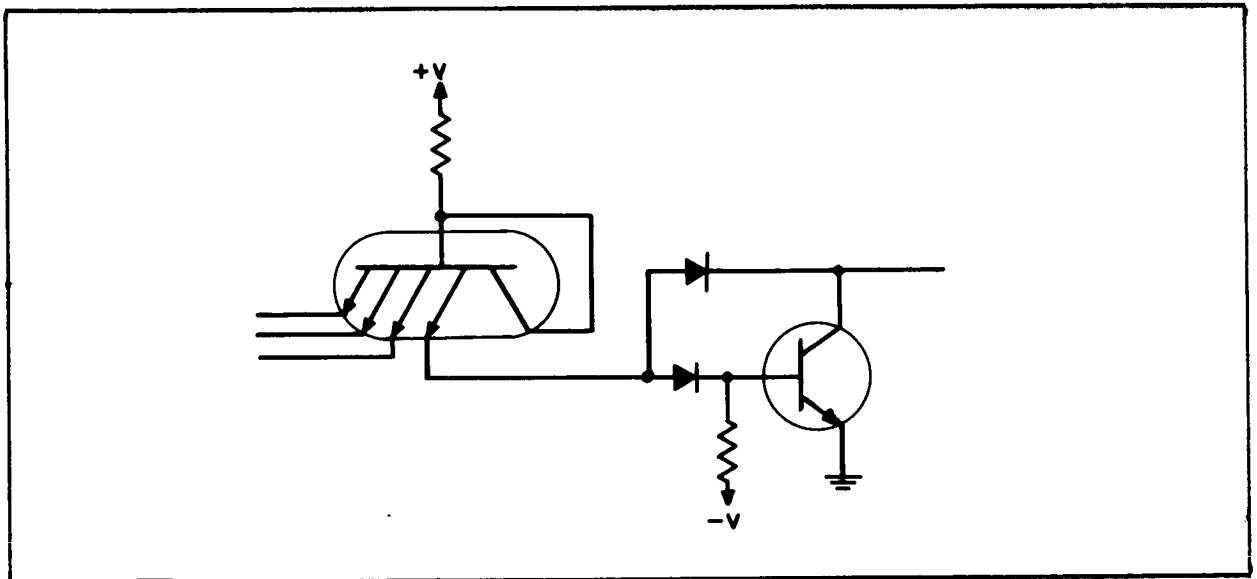


Figure 12 NS-D-TL Non-Saturated-Diode-Transistor Logic

Type III-A coupling, shown in figure 13, is a variation referred to as T-TL which uses transistor coupling to obtain improved response. Logic operation is equivalent to D-TL when inverse transistor gain (β_I) is low; coupling transistor action removes stored charge during turn-off, and generally permits the elimination of the output transistor base bias resistor.

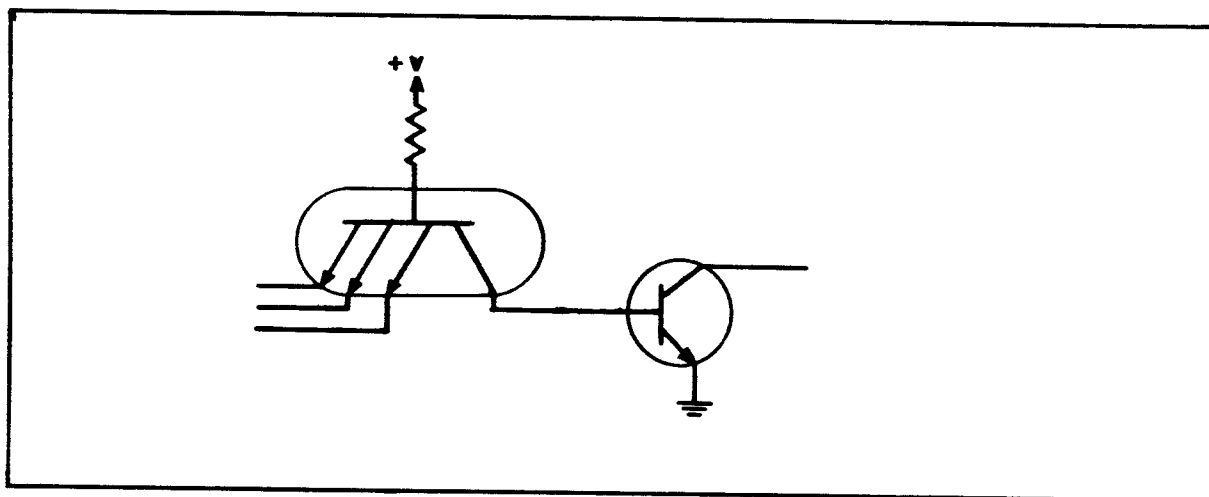


Figure 13 T-TL Transistor-Transistor Logic

C. Comparison of Logic Types

A comparison of the types of circuits described above is shown in the table below for five types which are commercially available. They are arranged in the table in increasing order of the number of equivalent components required for a 3-input universal gate function. A larger number of components generally increases fabrication complexity and increases

power dissipation. The general characteristics of these logic configurations are discussed and compared in the paragraphs following the table.

The isolation and speed-power rankings for the three saturated logic types were obtained from "The Changing Prospective in Microcircuits", Electronic Design, February 15, 1963, p. 56. This article describes the result of a study of different types of logic for single substances conducted by PSI. The author observes that no one logic type is superior to all others for every application, but rather that the characteristics of each type must be considered according to the particular over-all system requirements.

The isolation ranking is a qualitative measure of the input loading, the isolation between inputs, noise immunity, and variation of input loading with parameter changes, internal failures, and output loading. Logic types with the highest isolation are ranked first; those with lower isolation are ranked in increasing order. The non-saturated logic types are inserted into the original ranking by a comparison of their general characteristics with those of the three saturated logic types.

The speed-power ranking is a quantitative measure of the product of propagation delay and power dissipation of the different logic types when similar components and techniques are used in fabrication. This

characteristic varies considerably according to the design and technology used for the construction of actual circuits. Logic types with the lowest power-speed product are ranked first; those with higher power-speed products are ranked in increasing order. The non-saturating logic types are inserted into the ranking order indicated according to available data.

TABLE I COMPARATIVE RANKING OF AVAILABLE LOGIC TYPES

NAME	Function for + Logic	Type of Coupling	Number of Components	Speed- Power Ranking	Isolation Ranking
T-TL	NAND	III-A	3	1	4
D-TL	NAND	III	5	3	2
NS-D-TL	NAND	III	6	2	3
R-DC-TL	NOR	II-A	7	5	5
NS-DC-TL	NOR	II-B	9	4	1

D. Description of Logic Types

Resistor-transistor logic (R-TL) is a basic scheme for providing the NOR function for NPN positive logic. The resistors are used for linear input summing into the output transistor, which is normally biased off unless at least one input is present. The bias may be increased to provide either the inverse majority or the NAND output. The addition of speed-up capacitors to the input resistors, although significantly increasing transient response, is not sufficient to reduce the power-speed product to that available with other types of logic. The bilateral interconnection may create interaction problems between inputs; performance of the device is sensitive to variations of the input resistors, biasing, and transistor gain. The difficulty of fabricating an integrated resistor-capacitor combination for each input further decreases the suitability of this type of logic.

Direct coupled-transistor logic (DC-TL) is a theoretically simple method of performing the NOR function for NPN positive logic. Inputs are applied directly to transistor bases; the common collector is the output. Actual operation, however, is limited by the high sensitivity to parameter variations, input current "hogging" and low input impedance which limits fan-in and fan-out, and the low noise margin. These severe limitations have resulted in the actual use of a modified version (R-DC-TL) which includes a low impedance resistor-capacitor combination on each input to reduce the sensitivity to noise, parameter variations, and current "hogging". This modification increases power dissipation, propagation delay, and fabrication complexity. Since the fan-out capability of most NPN positive logic NOR

schemes is derived from the output collector resistor, the power dissipation must be increased to allow fan-out capability regardless of whether the fan-out is used or not.

The basic DC-TL scheme may be modified to provide non-saturated input logic (NS-DC-TL). The common emitter resistor reduces the problems of input current "hogging", and increases input impedance so that this type of logic offers high input isolation. Various methods may be used to provide outputs; both the OR and NOR may be provided conveniently. Good matching of components and close tolerance on a special reference voltage supply are required. The clocking function may be obtained by controlling the negative voltage supply by gating or a sinusoidal voltage. A two phase clock is required for flip-flop functions more complex than simple storage. An additional transistor, which shares a common collector with other input transistors, is required for each input. The voltage difference between the "1" and "0" level is usually very small, resulting in reduced DC stability and noise margin. NS-DC-TL offers high speed operation at the expense of high power dissipation.

Diode-transistor logic (D-TL) is probably the most popular type of integrated circuit logic, due to its similarity to discrete component circuitry and the excellent operating characteristics. D-TL circuitry operates with wide parameter variations to minimize the possibility of malfunction due to drift failure. Actual failure testing has shown that redundant D-TL is not sensitive to most catastrophic failures. D-TL is most commonly available as NPN positive logic NAND integrated circuits.

The newer versions of commercially available D-TL circuits offer about the lowest power-speed product available for circuits operating at moderate speeds and with good noise margins. Consideration of integrated circuit characteristics has significantly reduced the number of individual isolated components compared to the number of discrete components required for an equivalent circuit. The entire input diode array, as well as one level-shifting diode, may be constructed as one multiple-emitter transistor. Each additional input merely requires an additional emitter connection.

Transistor-transistor logic (T-TL) is a simplified variation of D-TL employing transistor coupling directly to the base of the output transistor. The elimination of one coupling diode reduces the noise margin and voltage swing to about the equivalent of DC-TL. Input isolation is similar to D-TL, except that inverse gain of the coupling transistor allows some "hogging" of input current. The inverse gain cannot be reduced without increasing the offset voltage of the coupling transistor*; increased offset voltage, in turn, decreases DC stability and noise margin. Increased speed at low power levels is possible because the coupling transistor removes stored charge from the output transistor to reduce turn-off time.

The output inverter of D-TL may be designed to prevent saturation to reduce excess drive and stored-charge effects. This may be accomplished by limiting the minimum "0" output voltage by a base to collector clamp to prevent saturation of the output transistor, as shown above for non-saturated diode-transistor logic (NS-D-TL). The increased "0" output voltage will, however, be more constant with increases in output loading,

$$* V_{CE(sat)} \approx \ln \alpha_I = \ln \frac{\beta_I - 1}{\beta_I}$$

if sufficient gain is available. Logic operation is equivalent to D-TL with increased speed and lower power dissipation under comparable conditions. Additional gain may be easily obtained for D-TL by substituting an emitter follower for the final level shifting diode.

The speed-power performance of some of the commonly available logic elements currently available are shown in figure 14. This figure shows the advertised performance characteristics of different logic types available from different suppliers.

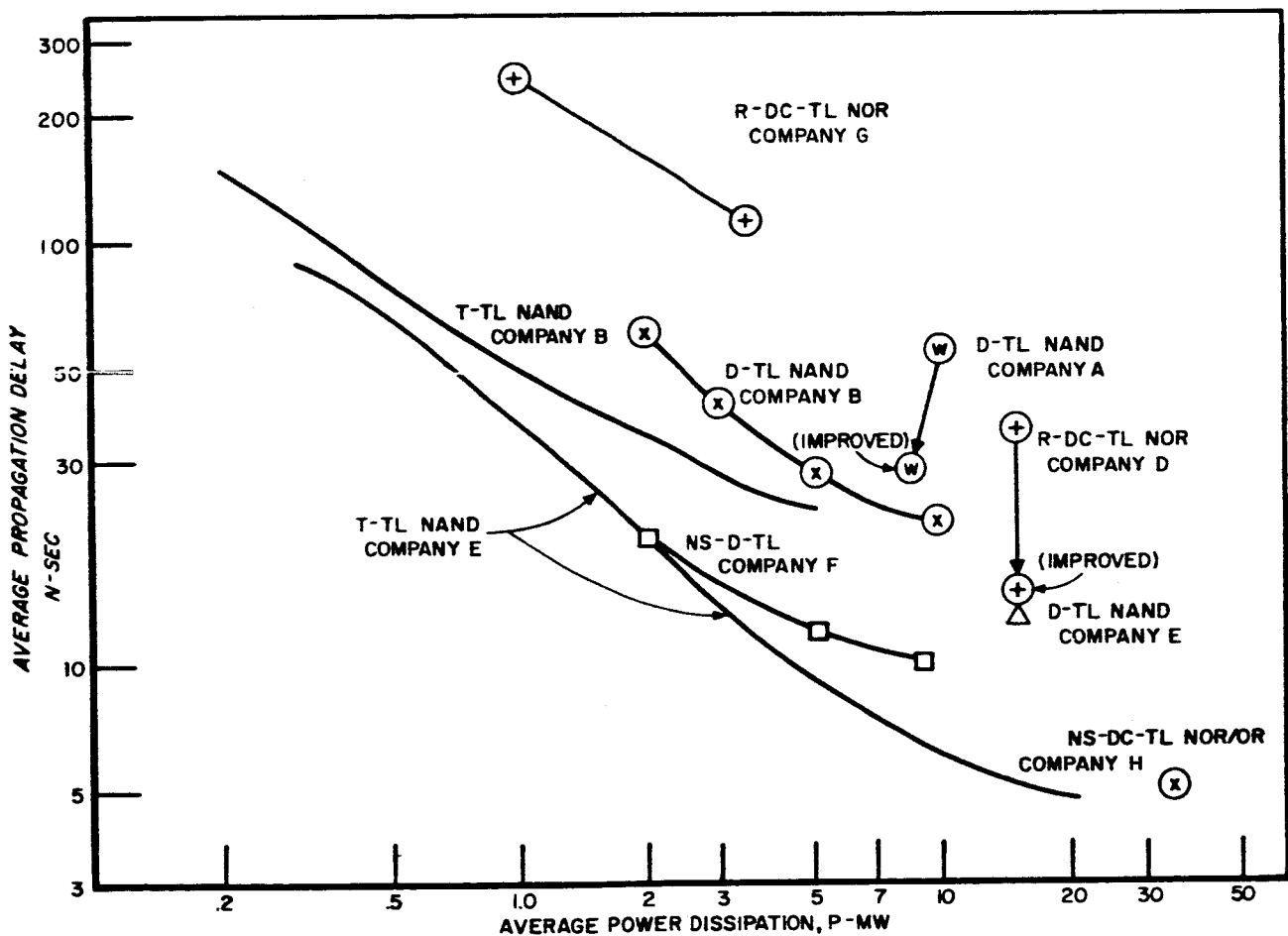


Figure 14 Speed-Power Performance

The wide variation of performance characteristics for different suppliers of the same logic types is due to several causes: differences of circuit parameter design, lack of standard test conditions (temperature, fan-out, voltages, etc.), as well as the rapidly improving technology in this field. Two recently announced improved versions of previous elements (Company A D-TL and Company D R-DC-TL) are indicated in the figure. The rapid rate at which improvements have been made in the field of integrated circuits makes it impractical to make an arbitrary decision to use only one logic element for all future spaceborne redundant systems. General characteristics, as well as the specific requirements of redundant systems, may be used to make recommendations, however, based on available information. The general characteristics discussed below may be used as a guide to the choice of circuits, even through exact requirements may vary.

Since systematic redundancy is most efficient and powerful when the basic elements are highly reliable, the realization of high system reliability with minimum weight and power penalties requires circuitry with high basic reliability. High circuit reliability, especially for extended periods of time, is usually realized when the circuit configuration is such that proper operation is not excessively sensitive to parameter variation or environmental extremes. High speed performance does not appear to be a particular requirement for most spaceborne systems; low power dissipation

is a much more desirable characteristic. Available power (and total energy) is often limited on space missions; the additional circuitry required to reduce the probability of system failure will further emphasize this problem. The power required by individual circuits must be held to a minimum to keep total power within available limits. The reliability performance of most integrated circuits depend on the temperature stress. The use of low power circuitry is an important factor in reducing the temperature stress, which, in turn, improves the basic reliability and performance characteristics of the individual elements.

Although T-TL offers high speed at low power levels, its sensitivity to parameter variation, noise, and input current "hogging" has reduced the general suitability of T-TL. This sensitivity appears to be a major disadvantage because the individual circuits in a redundant spaceborne system are required to operate reliably despite severe environmental variations and the occurrence of failures within the system. Since inverse transistor action can limit the input voltage signal, failures within the circuit or on the output may affect the inputs. This transfer of failure effects to inputs would be a serious disadvantage in redundant systems, where the effect of failures must be minimized.

DC-TL appears to be even more sensitive to parameter variations and failure effects, except for the various modifications which are used to reduce this problem. Positive NOR logic appears to be particularly vulnerable to output failures resulting in failure of input signals. This occurs because the transistor turn-on current is obtained from inputs; any

input must be able to provide sufficient drive to cause the output to be "0" for proper operation. Fan-out capability is obtained by providing each output with the ability to drive several inputs. If actual failures may cause all of the inputs to a circuit to be overloaded, then any other circuit receiving any of these inputs are also effectively failed. Additional fan-out capability is usually reflected in increased power consumption, which, in turn, increases reliability problems.

In contrast, the turn-on current for positive NAND logic is obtained within each logic element. This drive current is diverted to a low impedance input whenever any input is "0". Fan-out capability is provided by the output transistor gain, and may be increased without significantly increased power requirements. Since drive current is provided by each circuit, rather than by inputs, failures within an NAND circuit usually do not affect proper operation of inputs. The back-to-back diode coupling also offers good isolation characteristics. Actual failure testing has verified that failure effects in D-TL is usually limited to the circuit in which the failure occurs.

Limited testing for the effects of both transient effect of high gamma radiation and the permanent effect of integrated neutron flux has shown that D-TL integrated circuits are more resistant to radiation than forms of DC-TL.⁶ The transient effects of high gamma radiation appear to be primarily due to the leakage of the collector isolation diode. DC-TL is more susceptible because the larger number of common-collector transistors used creates a larger junction area. DC-TL was seriously affected at

gamma levels of 10^6 to 10^7 R/sec, while one company's D-TL withstood an order of magnitude increase. The same company's D-TL also showed more resistance to integrated neutron flux, but no microcircuits showed damage at ordinarily expected dosages. At a flux dose of 2.8×10^{14} neutrons/cm² (equivalent to about 100 years of continuous exposure in the Van Allen belts), one company's elements failed, another showed waveshape deterioration, while another microcircuit brand and discrete component D-TL showed no noticeable effects.

E. Logic Selection

Integrated D-TL circuitry appears to be the most appropriate type of logic for general use in redundant logic systems for spacecraft missions. It has been chosen for the general advantages of features described above, and particularly for its suitability for use in redundant spaceborne equipment, which requires both high immunity to noise and parameter variation, as well as reasonably low power dissipation. These requirements are generally not available in the various forms of DC-TL. Although T-TL logic is equivalent to D-TL, currently available elements are too sensitive to input current "hogging" to be suitable for use in redundant systems.

D-TL is known to have high noise immunity, good input-to-output isolation, good capability with other circuitry and relatively low power consumption. D-TL is particularly insensitive to drift failures; failure testing had shown that the effect of most catastrophic failures is not especially harmful in redundant logic networks. The speed capability of

available integrated D-TL circuits appears to exceed the requirements of most spaceborne systems. Some of this excess speed capability may be traded for lower power requirements by reducing the power supply voltages. Power dissipation could be further reduced by a redesign of present D-TL circuits to use higher resistance values. High resistance is a difficult problem in present circuits, since the characteristically low resistivity of diffused resistors requires a large area for high resistance values. The use of thin film resistors and capacitors on the silicon block in which the semiconductors are diffused, as planned by Westinghouse for the near future, would permit circuit design for significantly lower power dissipation without the large areas and narrow strip layout required for totally diffused circuitry. Such single-chip hybrid circuits are not presently available for general logic use.

It is expected that the positive logic NAND function will be used, since this permits logic design of functions as the sum of products, which is convenient for reduction and simplification by familiar methods. The NAND circuits shown are particularly versatile, since the collector outputs may be connected together to form AND-OR-NOT logic functions directly. R-S flip-flops may be formed by interconnected NAND elements; formation of more complex functions such as a compatible counter element require a large number of NAND elements and a two-phase clock. The majority voter is not a commercially available element, but it is easily constructed from NAND elements.

F. Majority Voter Design

Failure testing has shown that particular care must be used for the design of restoring elements so that failures on one input to the restorer do not cause failures on other inputs, and the failures in the restoring elements do not cause failure of a majority of inputs. This testing has shown that a conventional majority element (whether constructed as the minimum discrete component circuit, or of interconnected NOR or NAND elements) may experience failures which either cause immediate failure of the entire set of restorers, or which would cause the same result if a single input error occurs.⁷ If such effects are overlooked, the system reliability may be seriously degraded. Shown in figure 15 is a three input majority element using NAND elements which cannot cause an entire set of restorers to fail due to any single failures.

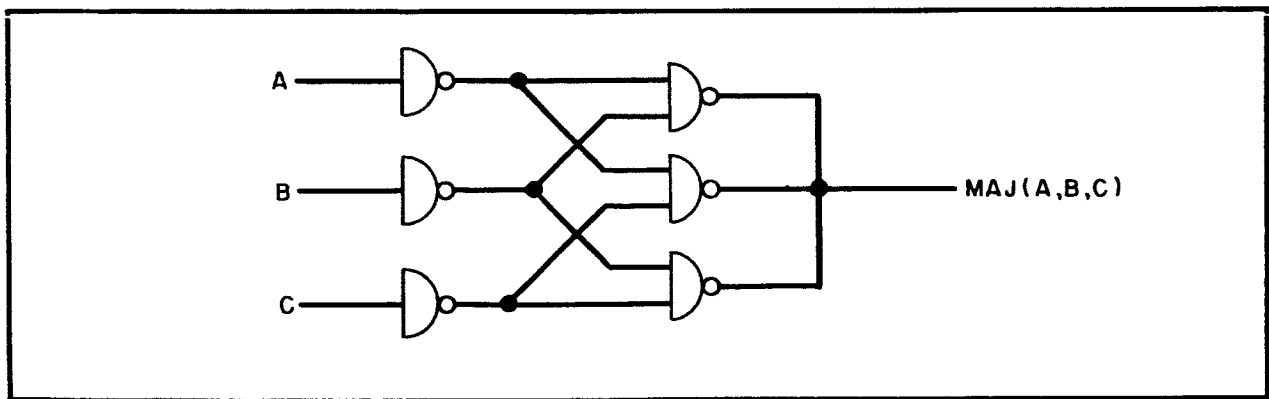


Figure 15 Majority Element with Input Isolation

The NAND implementation shown utilizes common output logic so that the voter requires only two more gates than conventional majority voters, and retains a two element input to output propagation delay. NOR implementation, however, would require a total of eight gates and four element input to output propagation delay to obtain input isolation for NPN positive logic. It is expected that the isolated input majority element shown will be more reliable in normal operation (all inputs alike) than a more conventional configuration, since very few single failure modes can cause the output to disagree with the inputs when all inputs are identical.

If higher orders of redundancy are used, then each input is provided with isolation gates. Since component redundancy is not used to protect against single failures, a simple test consisting of monitoring the logic output while applying all combinations of logic inputs will completely test the operation of the circuit. A custom-packaged majority voter would significantly reduce the size and weight of a redundant system when compared to one using individual packages. The packaging of this majority voter is of particular importance because it is used repetitively in a redundant system.

IV. Failure Testing of Redundant Systems

A. Introduction

1. Characteristics of Redundant Systems

The outstanding attribute of a redundant system is that of providing high reliability for a longer period of time than the non-redundant counterpart. Typical reliability curves depicting this relationship for a simple system shown in figure 16. It is assumed here that both systems begin operation with all circuits, subsystems, wiring, etc. in a failure free condition.

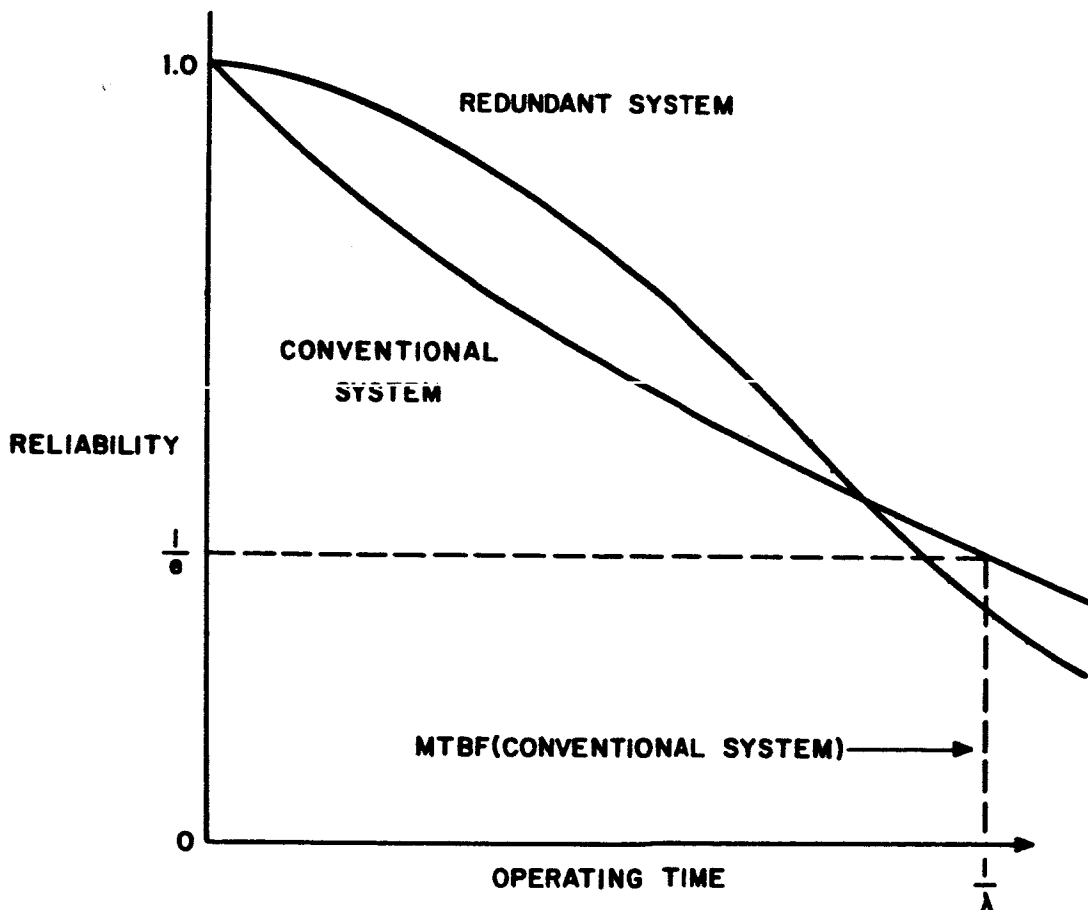


Figure 16 Reliability of Conventional vs. Redundant Systems

The statistical relationship between reliability and operating time is derived by assuming that failures occur at constant rate and are inherently random and independent. After some period of operation without maintenance, the reliability of a typical multiple line, majority voted redundant system falls off and becomes less reliable than the non-redundant version. This behavior is normal since the greater number of components subject to statistical failure eventually cause the majority voters to have incorrect outputs. The initially flat portion of the redundant system reliability curve is the characteristic which is exploited to provide high mission reliability.

Since current spaceborne equipment is unattended after mission commencement, it is important to assure that the equipment is in perfect working order "before launch". It may not always be practical to completely test each part of a redundant system after final assembly and installation into a space vehicle, and thus the term "before launch" includes diagnostic testing before final assembly. It will be shown that a redundant system may be conveniently diagnosed for the presence of failures after final assembly and installation in a space vehicle. This may be accomplished during the pre-launch test period when the vehicle is about to begin its mission. Essentially the technique employed is that of removing the failure masking effects of redundancy and testing the replicated systems separately.

The function of these tests is initially to detect the occurrence of a failure and secondly to determine its location. The tests would be

useful in deciding whether the equipment should be finally assembled and installed into the space vehicle or if the equipment is free of failures and ready for launch. The goal here is to assure that all of the initial failure protection which has been designed into the system is available.

In a non-redundant system the best one can do is to test the system and then hope that no failures occur. The statistical nature of failure occurrence, however, offers little assurance that a failure will not occur just after mission commencement. This occurrence often precipitates total mission failure in a non-redundant system. The redundant counterpart is obviously better suited to tolerate random failures. Further, a typical order three redundant system which has been diagnosed to be free of failures prior to mission commencement is not vulnerable to single failures and thus offers a high degree of assurance of mission success.

Further tests would be utilized to isolate and locate the failure. The goal here is to effect repair and thus return the system to perfect working order. Since this may consume considerable time and involve special repair or replacement facilities, a duplicate system, which has been found free from failure, may be required to expedite scheduled installation into the space vehicle.

For redundant systems which receive maintenance the purpose of diagnostic testing is again to detect and locate failures. The goal, however, is to return the system to perfect working order and thus assure the highest possible reliability during the entire operational life of the equipment. In order for periodic maintenance to be effective it follows that the

period between maintenance checks should be sufficiently short so that the reliability for the maintenance period is high. The probability of operation repeatedly traverses the initially flat portion of the redundant reliability curve.

The general problem of diagnostic testing is to provide suitable test facilities and methods which are effective in determining whether a failure has occurred, and to determine its location. In a redundant system the implementation of test facilities entails many considerations, ranging from basic system configuration to the details of circuit design. In a conventional non-redundant system, test provisions are all too often given only token consideration. Although the test features provided may be ineffective or inconvenient, the diagnosis, failure location and repair of the equipment is often made possible through the ingenuity of an experienced technician. A redundant system similarly encumbered imposes a much more difficult task. Thus the need for integrating system configuration and test facilities in the initial design stages becomes extremely important.

2. Testing of Conventional Systems

The techniques for detecting a failure in a redundant system represents a problem which is alien to the test philosophy of conventional systems. In a non-redundant system the effect of a failure is rather dramatic and is usually evidenced by either partial or total system failure, or obvious changes in operational behavior. This simplifies the problem of detecting an error, but is small consolation to the user who loses the service of a system without warning, perhaps at some crucial moment. Total

system failure usually indicates the failure of a major function, such as a power supply or clock generator. Changes in operational behavior and partial failures normally provide symptoms which, when analyzed, are valuable in converging on the failure location. In a redundant system the effect of a non-critical failure is not evidenced by any change in system behavior. This means that the effect of a failure does not provide gross symptoms which may be used to indicate its occurrence or determine its location. The solution to this unique problem is suggested through several avenues of approach which represent diagnostic routines and implementation schemes unique to redundant systems.

Before considering the unique demands which a redundant system imposes on the required test facilities, it is useful to consider some approaches which are applicable to digital systems in general. These general approaches include waveshape monitoring and the application of various stresses to enhance the chance of detecting present or potential failures. The combination of general approaches with the specific approaches to be suggested appear to offer a more inclusive repertoire of techniques from which to choose.

In a conventional system a failure of some circuit or sub-system normally provides an indication of its occurrence by the resultant changes in operational behavior. These are usually designated as catastrophic failures. Degraded components which are not sufficiently marginal to cause circuit failure are more difficult to detect because there is no indication of a change in system behavior. Often, however, a degraded component may

be detected at the circuit test point level by changes in normal wave-shape. At the component level the degradation may be considered as a failure. At the circuit level this condition represents an impending failure. Understandably it is important to detect and repair impending failures since it is very likely that the circuit will soon fail. This is one of the more important aspects of periodic maintenance of non-redundant systems. Often the system may be operated normally and the various test points monitored to detect marginal voltages, wave shapes or rise times. This represents a very time consuming procedure and is severely limited in effectiveness by the number of test points which are provided. Many marginal components are then essentially undetectable.

Another problem which often arises is when a failure in circuit operation becomes sporadic. In this case the system may operate normally for most of the time making the location of the fault a difficult task. As so often happens, just as maintenance personnel are in the process of converging on the fault location, the fault disappears and the system operates normally. The problem here is that the fault is not present long enough to allow an adequate diagnosis of the difficulty.

A more powerful approach for locating impending and sporadic failures involves the application of stress to the system. This will often precipitate a circuit failure by subjecting components to a condition which magnifies any degradation. Consider now the two general classes of approaches for imposing system stress--environmental and electrical. Environmental

stress may be typically sub-divided into temperature, humidity, pressure vibration, shock, radiation, etc. The application of one or combination of these environmental stresses is seen to present three main problems; 1) the size, complexity and cost of the facilities required, 2) the difficulty of performing measurements in an alien and often dangerous environment, and 3) the possibility of subjecting components to unnecessary stresses and thus causing unwarranted damage or destruction.

Temperature stress is perhaps the most popular approach because of its utility in causing parameter changes in resistance, capacitance, leakage, gain, threshold, etc. A second advantage is the small amount of additional facilities which are required. Often, temperature stress may be conveniently applied by controlling the system cooling to increase or decrease operational temperature. Component variations caused by temperature stress often make circuit operation marginal when such changes are beyond the normal specified design limits. Thus a component which has become only slightly marginal at normal operating temperature, and is indicative of impending failure, may be magnified by temperature stress to precipitate circuit failure. This method is often used, for example, in testing transistors for leakage current degradation at elevated temperatures. In a system test the increased leakage current of degraded transistors causes circuits to become sufficiently marginal to effect circuit failure.

The remaining types of environmental stress are difficult to impose on a system without test facilities of vast complexity. For this reason

they are not readily amenable to system testing but find greater utility at the component or sub-system level. A case in point is the development of highly reliable components, i.e., by carefully controlled production followed by extensive testing under a variety of environmental and electrical conditions.

Electrical stress is a more convenient method for detecting marginal components and impending failures. A convenient method for stressing an entire system simultaneously is that of marginal voltage testing. In this approach the system power supply voltages are varied to combinations of maximum and minimum levels for which the circuits were designed. When all defective components, modules or sub-systems have been detected and replaced the system power supplies are returned to their nominal values. Marginal voltage testing is often combined with simulation routines and static and dynamic measuring techniques to provide an inclusive test program.

Simulation programs provide a form of electrical stress which is seen to exercise the variety of operational functions which a system may be required to perform under actual operating conditions. Often however, a simulation technique may subject the system to operational speeds which are not encountered in normal system operation. This might be accomplished by varying the frequency of system clock generators to either increase or decrease the speed of operation. In a spaceborne sequencer, for example, it may be necessary to speed up the occurrence of time events by several orders of magnitude in order to test all functions in some reasonable test period. In other applications increasing the speed of operations to the

maximum design limit is often useful for magnifying the effect of marginal components. For example this technique is seen to be useful in determining degradation in capacitive coupling circuits.

A reduction in operating speed does not usually subject the system to stress but is useful in ascertaining that some normally fast sequence of operations is being performed correctly. Here, the reduction of clock rate is utilized to allow operation sequence to be conveniently monitored. The general approaches discussed are primarily useful in precipitating static failures which are impending or sporadic. DC failures and catastrophic failures are usually immediately apparent from the manner in which the system behaves. When only a portion of the system fails in the static state it often provides symptoms which may be used in diagnosing the location of the failure. If a failure occurs near the "front end" of a system, the majority of outputs will usually become static. In this case the symptoms are not sufficiently explicit to allow an adequate diagnosis. Simulation equipment then becomes useful in determining the failure location. This is accomplished by applying suitable signals at the various subsystem inputs and monitoring outputs for the presence of the correct response.

3. Failure Detection in Redundant Systems

The problem of detecting a failure in a redundant system is usually more difficult than in the conventional counterpart, because the effect of non-critical failures do not provide gross symptoms of their occurrence. This difficulty in diagnosing a failure is amply compensated

by the vast improvement in reliability which a redundant system provides.

Since a conventional system normally provides little indication of an impending failure, the only available resort by which the system quality may be diagnosed is by the application of stress. It is, however, an inconclusive test of the systems ability to perform reliably. In a redundant system the application of stress to components and circuits for the purpose of detecting impending failures is not of significant value because the effects of individual failures are masked by the system configuration. Although redundant systems are able to tolerate failures without causing total system failure, it is often desirable to diagnose the system to detect any internal failures. It will be shown that the application of conditions which reduce the ability of a redundant system to withstand internal failure acts like stress by modifying the configuration so that the failure masking effects are removed. In this manner, failures which are present will be indicated by the behavior of the system. The following paragraphs will describe techniques for detecting and locating failures in redundant systems.

An order-three, multiple-line, majority-voted redundant shift register system will be used to demonstrate basic approaches. This is done for ease of explanation and is not intended to suggest that the approaches may not be extended directly to more general system configurations, or to higher-order redundant systems. It may be noted that the testing of redundant systems will involve a hierarchy of tests involved with first testing the signal processing parts, then the testing of the restoring elements, and finally the testing of the hardware added for the initial testing function

: itself. The extent and complexity of this hierarchy will depend on the confidence which is required of the tests and the degree of automation desired. It appears impossible, however, that perfectly reliable operation can ever be expected from any hierarchy of imperfect equipment monitoring other equipment. Although these testing methods are intended to make a significant contribution to the techniques available for testing redundant equipment, it is expected that further work in this area will result in further improvements. The accuracy and complexity of the tests should be balanced to obtain efficient system operation.

Often, the problem of failure detection is directly connected with the requirement for determining the location to facilitate maintenance repairs. Therefore, some of the more complete testing methods will include combined detection and location. Although failure location techniques are usually more complex than the basic failure detection techniques they often include complete failure detection capability in order to locate all failures which might exist in a redundant system. Failure location techniques also provide effective methods to detect and locate failures in the failure detection and location circuitry itself.

Basic failure detection will probably be most useful as a verification technique to indicate that at least a major portion of a redundant system is failure free. This will assure that the failure protection which has been designed into a redundant system is available to prevent system failure. Simple failure detection techniques are also expected to be a preliminary technique which will indicate if any failures are

present in a maintained redundant system, so that further corrective action may be undertaken. It is important that all failures be detectable in a maintained redundant system, so that failures are not allowed to accumulate and degrade system reliability.

4. Failure Location in Redundant Systems

If a failure is known to exist in a redundant system, it is often desirable to obtain further information concerning the location of the failure. This is generally required so that the module containing the failure may be repaired or replaced. Although it is very desirable to be able to detect any failure to permit maintenance, it is only necessary to locate failures to within the smallest replaceable module. Therefore, the requirements of failure detection depend strongly on the contents of the smallest replaceable module. If entire subsystems are contained in a module, then each subsystem could be provided with independent failure detection hardware. This would be sufficient to locate failures within the replaceable module. It is possible that the requirement for test points at each replaceable module to permit failure location may in turn determine the practical size and contents of the module. If the test points and connections occupy a large space compared to the basic module, then the volume efficiency is rather poor, and a larger replaceable module might be more practical.

If repairs are expected to be made while the system remains in operation, then the module which contains the failure must not include the remaining replications of that function. This is necessary to permit the system to operate while the module containing the failure is removed.

If the entire module is to be replaced if it contains a failure, then the failure location technique must be sufficiently accurate to determine which module contains the failure. This module may then be replaced without interruption of normal system operation. Maintained redundant systems which are continuously monitored and repaired require a combined failure detection and location technique which may be applied without altering the operational characteristics of the system. It will be shown that relatively complete testing may be accomplished during system operation. This is possible because the most frequent and harmful failures usually cause signal disagreements at the inputs to the voters. These signals may then be compared, either automatically or with the use of test points, to detect and locate these failures. Certain system configurations are amenable to controls which allow complete failure detection and location with access only to the signals at the inputs to the voters. More generally applicable techniques require access both to the voter inputs and outputs. These techniques, as well as the implementation circuitry required, are described in the following paragraphs.

5. Signal Comparison in Maintained Systems

The location of a failure in a conventional system requires that a handbook be provided to indicate the correct wave shape and binary sequence to be expected at each location. This is in addition to simulation equipment which may be required to place portions of the system into dynamic operation. The redundant system masks the effect of individual failures and thereby makes the task of detecting their occurrence more difficult. It will be shown, however, that the masking effects of a

a redundant configuration may be conveniently removed by controlling the outputs of the signal processors. This is essentially a gross system approach whereby the occurrence of a failure is indicated by forcing the system to assume various vulnerable configurations. If the system is allowed to either operate normally, or in some configuration for which all operations are performed correctly, the detection and location of failures may be conveniently accomplished by examining replicated elements for signal disagreement.

In many respects, the location of failures in a redundant system is a much easier task than in the conventional system counterpart. This is because an improper signal may be determined by comparison with its replicated versions. If a redundant system is operating correctly in an overall system sense, then the correct signal of each monitored element is available at least at a majority of associated test points. This is seen to eliminate the tedious task of monitoring elaborate wave shapes and sequences. Maintenance personnel are then presented with a system which, in principle, contains an integral handbook of normal signals to be expected at the various locations. The system may be permitted to operate normally, without simulation equipment, performing operations whose binary sequence at any single location is so complex that one could not hope to describe them adequately in any handbook. This suggests the possibility that maintenance personnel need not be completely familiar with the detailed operation of the system.

The determination of an error could be provided by a difference detector in combination with a suitable indicator. A technician would be required only to monitor the various test points in some prescribed sequence until arriving at the location of a signal disagreement. He would not be required to possess any special knowledge of what constitutes a correct or incorrect wave shape, binary sequence or repetition rate. Also, most difference detector devices which might be employed will signal any large departure from normal signals, and may include memory to indicate the location of transient or sporadic failures. From this we may conclude that the training requirements for maintenance personnel may be appreciably reduced, thus providing redundant systems with a distinct maintenance cost advantage over the more conventional counterpart. This attribute alone might become a significant factor in evaluating the total utility of a redundant system which is periodically maintained.

In order to reduce the total system failure rate, periodic maintenance must be conducted at a sufficiently short interval so that individual failures are not so probable that system reliability is appreciably degraded. In addition, if system failure occurs it might be necessary to employ simulation equipment to place portions of the system back into operation. The advantage of not requiring simulation equipment to locate individual failures is an important feature of a maintained redundant system. Thus the function of periodic maintenance is not only to assure high system reliability during the life of the equipment, but also to eliminate the requirement for simulation equipment to locate failures.

Thus far in our discussion of maintained redundant systems, it has been implied that the signal comparison equipment is usually externally applied to the appropriate test points in much the same manner as an

oscilloscope or voltmeter is used in a conventional system. As indicated previously, it may be undesirable to provide these test points at every signal processor and voter output in the system. This may be due to the lack of access to the signals, the physical size of the test points in comparison to the circuitry being monitored, or the signal loading caused by test point leads. In some applications it may therefore be desirable to provide error detection and display as an integral part of the system. Integral signal comparators may be desirable for example, in a maintained redundant system which is continuously monitored during operation and each failure is repaired as soon as it is detected. This maintenance philosophy allows a much higher system reliability than available with periodic maintenance. With proper design it appears feasible to remove and replace defective modules without disturbing the operation of the system.

Since signal comparators will indicate only when signal disagreement occurs during the normal system operation, more extensive tests are required to detect and locate such failures as might occur in signal processors which are not to be used for some modes of system operation, some of the failures in voters, and failures that might occur in the control and signal comparison circuitry. This suggests a maintenance philosophy of continuous monitoring combined with periodic complete testing as follows: Signal processor outputs are continuously monitored during the operation of the system for the indication of the more frequent and harmful failures which cause incorrect signals. These failures are located and may be repaired without interrupting normal system operation. Periodically the normal

operation of the system is shut down to allow the system to be completely exercised and the otherwise undetectable failures to be located and repaired. In contrast, the periodically maintained system is allowed to accumulate failures, even though they may be easily detectable, until the end of a scheduled maintenance period. Continuous monitoring and repairing is therefore a very powerful technique for detecting and repairing most failures as they occur, without seriously impairing the ability of the system to operate continuously while individual failures are repaired.

B. Singular Rank Testing

1. Detection of Signal Processor Failures

An obvious method for detecting failures in a typical redundant system is to separate and reconnect the replicated parts to create individual, independent systems. Each system may then be separately diagnosed for the presence of failures in the conventional manner. This would require that the basic system be provided with a large number of special switching circuits which accomplish a separation. Such an approach is somewhat impractical because of the expense, complexity and reliability degradation which the additional circuitry and wiring would impose. As will be shown, a much simpler means is available to provide a pseudo-separation of replicated systems without requiring an elaborate switching mechanization.

As an example, consider the simple redundant configuration shown in figure 17. Each of the complete replications of the non-redundant system are hereafter referred to as a rank of the system. Each rank normally

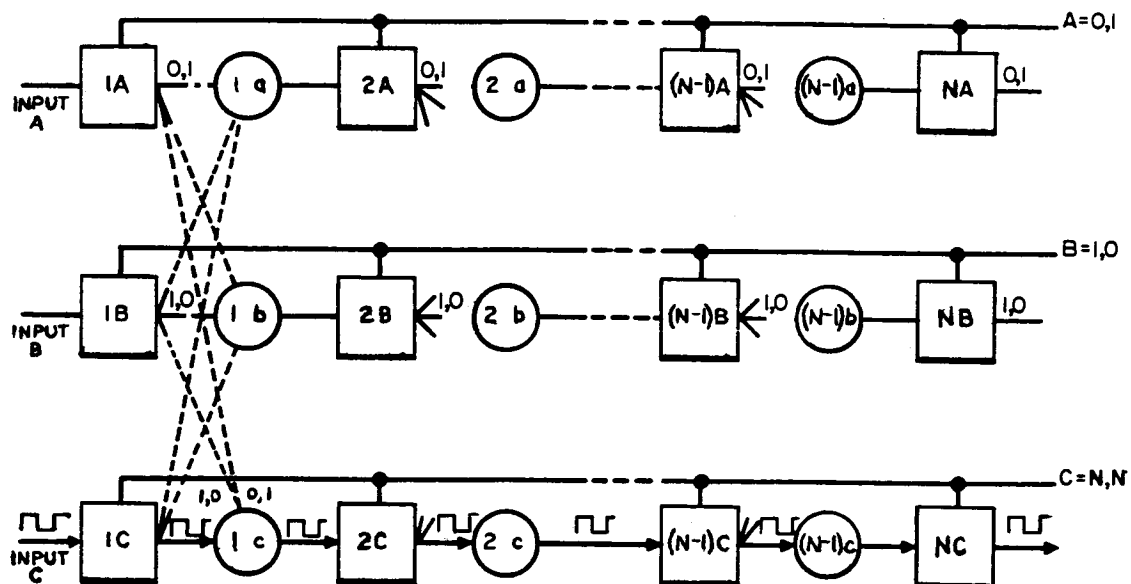


Figure 17 Singular Rank Testing

consists of the components of the non-redundant equivalent system, separated by the majority-voting restorers. Each of the signal processing elements (indicated by blocks) within the same rank are designated with the same capital letters; each of the majority voting restorers (indicated by circles) within the same rank are designated with the same lower case letters.

The corresponding replications of the same signal processors are hereafter referred to as being on the same file of the system. Each element in the file normally performs the same function, and is designated with the same number. Each signal processor file corresponds to individual functions at the non-redundant system. If a signal processor file has a restoring file associated with it, the restoring file may be assigned the same number.

It will be assumed that the order of redundancy is uniform throughout the portion of the system which is being tested and that the only interconnections between ranks occur at the inputs to restorers. Singular rank testing will assume that there is no restrictions on system size, configuration, or uniformity of direction of signal flow. These characteristics are chosen to be compatible with current redundancy synthesis techniques.

Suppose that the control lines shown in figure 17 provide a means of causing each output of the rank signal processors to assume either the "1" state, the "0" state or "N" (normal operation). In effect, the output of the A and B rank blocks have been forced to assume definite DC failure states. The mechanization to accomplish this is described in part D of this section, and will be shown to entail only slight modification to the normal circuitry. Consider the effect of causing all the A and B rank signal processors to assume a static complimentary state, allowing the C rank signal processors to operate normally, and that the system is allowed to operate with its normal inputs. Under the conditions that all A and B blocks are in a complimentary state the input to each voter consists of "1", "0" and the output of the preceding C rank signal processor output. This means that the dynamic signal predominates and causes this signal to appear at the output of the voters. If all voters operate correctly, the system is equivalent to a non-redundant system, and may be completely exercised in the same manner as the non-redundant system to verify that all signal processing blocks in rank C are functioning correctly. This test should also yield identical results if the

complimentary states of the A and B rank blocks are reversed. If an incorrect final output results for both tests it indicates that at least one failure is present in the C signal processors, the c voters or combinations of both. If only one test is successful, then a failure is evidently present in one or more of the c voters.

Success of either of the above tests is sufficient to verify that all C rank signal processors are failure free. It should be noted that the presence of a correct output for both complimentary test conditions does not verify with certainty that the c voters are failure free. This is because each voter was subjected to less than the maximum possible number of input signal combinations. Consider the various combinations of input signals and the correct response of a three input majority voter in the table below. States 1 and 2 represent the case when A="1", B="0", and C="N"; states 3 and 4 represent the case when the static signals on A and B are reversed. All signals are the same for states 5 and 6. States 7 and 8 occur when C disagrees with the other two inputs.

<u>State No.</u>	<u>A</u>	<u>B</u>	<u>C</u>	<u>Output</u>
1)	1	0	1	1
2)	1	0	0	0
3)	0	1	1	1
4)	0	1	0	0
5)	0	0	0	0
6)	1	1	1	1
7)	1	1	0	1
8)	0	0	1	0

Only the first four of the eight combinations were verified by the test conditions described. States 5 and 6 are trivial however, since they contain the combinational states of 2, 4 and 1, 3 respectively. If a majority voter makes a "1" output decision for inputs consisting of two "1"'s and a "0", it will make the same decision for an input of three "1"'s. Similarly, if a majority voter makes a "0" output decision for inputs consisting of two "0"'s and a "1", it will make the same decision for an input of three "0"'s. From this it appears reasonable to assume that if the majority voter operates correctly for the first four states it will operate correctly for states 5 and 6. Thus the combinations which have not been tested and hence explicitly verified are states 7 and 8.

The tests conducted thus far have verified that all C rank blocks operate correctly and that the voters operate correctly for six of the eight possible input signal conditions. The A and B ranks may be similarly tested with the result that the correct operation of all signal processing blocks may be verified. This test philosophy is seen to be an approach for isolating each rank of a multiple line configuration and thus determining the presence of any failures which would jeopardize the ability of the system to mask out future failures. Each rank is not operated simultaneously and independently, but rather one rank at a time is effectively removed from the multiple line configuration and separately diagnosed for the presence of failures.

The success of all of these tests has verified the proper operation of all signal processors. These tests have not completely verified the

condition of the voters as was described by the example of the C rank tests. However, the following voter input-output operation has been verified with certainty: All voters will make correct decisions if the input from the rank in which the voter is located agrees with at least one of the other inputs.

The condition which has not been verified is the uncertainty that a voter will make a correct decision when the input from the rank in which the voter is located is in disagreement with the majority of the remaining inputs (both remaining inputs for order three redundancy). It should be noted, however, that the complete set of singular rank tests will result in the application of all possible combinations of inputs to the voters. These tests are therefore sufficient to verify that any undetectable voter failures cannot combine with further single failures to cause an order three system to fail.

There are, however, a very limited number of component failures which can occur in the majority voter which cannot be detected with singular rank testing. These involve the failure of two of the input diodes for the three input D-TL voter. If the voter has a conventional minimum design, singular rank testing will indicate if either of these diodes is shorted. Due to the additional input isolation, the occurrence of these input diode shorts cannot be detected in the isolated input voter which has been shown in figure 15. If either of these undetectable diode shorts has occurred in the isolated input voter, the result is that the voter output is a "1" whenever the input from the rank in which the voter is located is a "1". The majority function is performed for all other inputs. The occurrence of either one of these

diodes being open cannot be detected for either the minimal design or the isolated input voters. The result of this condition is that the output of the isolated input voter is "0" whenever the input from the rank in which the voter is located is a "0"; if the input to a minimal design voter is a "1", the voter output is a "1". If one of the diodes shorts and the other opens, then the voter output is controlled by the input from the rank in which the voter is located, although the diode short could be detected if the minimal design voter is used. Therefore the existence of undetectable failures cannot introduce additional errors, but may cause signal processor errors to propagate through the restorers.

The above analysis has shown that the occurrence of undetectable failures tends to cause the output of the voter to be dominated by the signal from the rank in which it is located. In the worst possible case (complete dominance caused by the one diode open and the other diode short in every voter in every restoring file when these failures are undetectable), the restorers have been effectively replaced by conductive paths from the output signal processor in the previous file to the input of each following signal processors in the same rank. The result is equivalent to eliminating the restoring file completely (except that the reliability of the signal processors is reduced by the additional voter circuitry). Although it is extremely improbable that such conditions would predominate in a system recently constructed from completely tested parts, the system becomes more vulnerable to further failures if they are allowed to accumulate.

2. Detection and Location of Voter Failures

It may be desirable to have some means for detecting the presence of any failures within the system. One such example in which some method of complete testing is desirable is a maintained system which is expected to operate reliably for extended periods of time. If such a method is convenient, signal comparison may be combined with singular rank testing to detect and locate all voter failures. Since the combined singular rank tests result in the application of all possible inputs to the voter, the outputs of all voters in a restoring file may be compared for agreement while the inputs are applied. All voters are failure free if no output disagreements occur while all combinations of input signals are applied.

Since the only purpose of reversing the complementary states of the two ranks not being tested in an order three system was to gain additional information concerning the voters, voter comparison testing eliminates the need for interchanging the complementary states associated with each rank test. This requires, however, that a systematic method be used to assure that the complete set of tests results in the application of all possible combination of inputs to the voters, except the trivial cases when all inputs are the same. This condition will be met if the following rule is followed during singular rank testing: As each of the ranks is completely exercised as an individual non-redundant system, the particular pair of complementary DC states of the remaining two signal processors is chosen so that the state of either rank does not duplicate the DC state during any previous testing of the other ranks. Since the choice of which pair of

complementary DC states for the testing of the first rank is arbitrary, either of two alternate sequences may be used for the complementary DC states; these states will be complements of those in the alternate sequence. Thus it may be shown that only three tests (one for each rank) are required for complete singular rank testing with signal comparison. If each test is successful in demonstrating that the system will perform the entire set of functions for which it was designed, all signal processors are verified to be failure free and the voters are capable of transmitting a correct dynamic signal for some of the possible input states. If, in addition, all voters make the same decision while the proper sequence of controls is applied during the above tests, the voters are verified to be failure free.

3. Detection and Location of Control and Comparator Failures

The basic concepts of singular rank testing may be extended to verifying that the controls used for singular rank testing are operating correctly. Rather than allowing each rank to operate individually, each rank is individually controlled by the singular rank testing controls. If the controls are working properly, a signal comparison on the output of each signal processing file should indicate a disagreement whenever the dynamic signal on the remaining ranks is in disagreement with the DC state of the rank being controlled. In the case where difference detectors are used on the output of all signal processor files, this testing will also test these difference detectors. The detectors should indicate a difference at each signal processor file whenever the signal on the controlled rank disagrees with the dynamic signals. If the signal comparison of the signal

processors is accomplished while complementary DC states are applied to each pair of ranks, as described above, all possible input combinations involving disagreements are applied, and the difference detectors should give a continuous indication. If signal disagreements are noted for each signal processing file while all of the ranks are being controlled (either individually, in pairs, or for all possible input combinations involving disagreements, but not when the entire system is allowed to operate without signal processor failures) then the associated singular rank control circuitry is verified to be failure free.

4. Summary

It may be concluded that singular rank testing techniques are a very powerful tool for verifying that a redundant system does not contain internal failures. This testing would be valuable for use in acceptance tests which verify that all the reliability designed into a redundant system is available, or as the failure testing for continuously monitored and repaired systems with periodic complete verification, or in a system which is only periodically diagnosed to determine if any repairs are needed. The basic singular rank testing is a simple and effective method to allow a redundant system to be tested as if it were a non-redundant system to verify that all signal processors are operating correctly, and that the restorers will introduce no additional errors. This is equivalent to verifying that an order three system is not vulnerable to single failures. Basic singular rank testing techniques may combine with signal comparison to detect and locate failures which may exist in the signal processors, the restorers, the

control equipment, and any signal processor difference detectors.

Failure detection and location are often directly associated problems; failure location techniques are also effective failure detection techniques when they are available. It is expected that basic singular rank testing will be used as an effective and efficient technique for verifying that a redundant system is nearly failure free for regularly scheduled maintenance, or for relatively simple acceptance tests. The more complete detection and location techniques are expected to be used for the more thorough maintenance checks where any failures would be repaired, or for complete final tests after assembly. Signal comparison on all signal processor outputs may be used to continuously monitor and locate most failures in a continuously maintained system. These tests can be designed as part of almost any majority voted, multiple line system with a uniform order of redundancy throughout the portion being tested. No special signal simulation equipment is required, except the normally required inputs. The equipment required for the tests is described in more detail in part D of this section.

C. Interwoven Rank Testing

1. Complete Failure Detection

In some systems it may be desirable to completely diagnose a redundant system without the use of the signal comparison and failure location technique described above. In some cases, it is possible to perform this diagnosis without the requirement for any of the test points necessary for signal comparison. One such technique, which will be described

in the following paragraphs, is referred to as interwoven rank testing. It represents an extension of the singular rank testing, since the signal paths are interwoven between the ranks to form an equivalent non-redundant system in which the signal is switched from one rank to another at the restoring files. This is possible only if the system configuration has a sufficient degree of regularity. The example will assume that the system has restorers on the output of every signal processing file, and that these files may be assigned odd and even numbers in such a manner that odd files receive inputs only from even files, and likewise that even files receive inputs only from odd files. These restrictions are in addition to the assumptions on which singular rank testing is based. It will also be shown that the controls used for failure detection may be used to locate voter failures without requiring test points or difference detectors on the output of the voters. Comparison of signal processor outputs is sufficient to continually monitor signal processors and locate all voter failures.

Shown in figures 18 and 19 are six replications of the previously discussed configuration, with the exception that the two control lines for each rank individually determine the state of the odd and even numbered signal processors. If the two control lines for each rank were connected, the system would be identical to the one used in describing singular rank testing. Consider that the control lines and associated signal processors are placed in the following states: AO="0", AE="1", BO="N", BE="0", CO="1", CE="N", as shown in figure 18a. If an input signal is applied to the first file of signal processors, the signal flow will take the path shown by the arrows. This is because the two remaining signal processors in each file have been placed in complimentary static states. If all signal

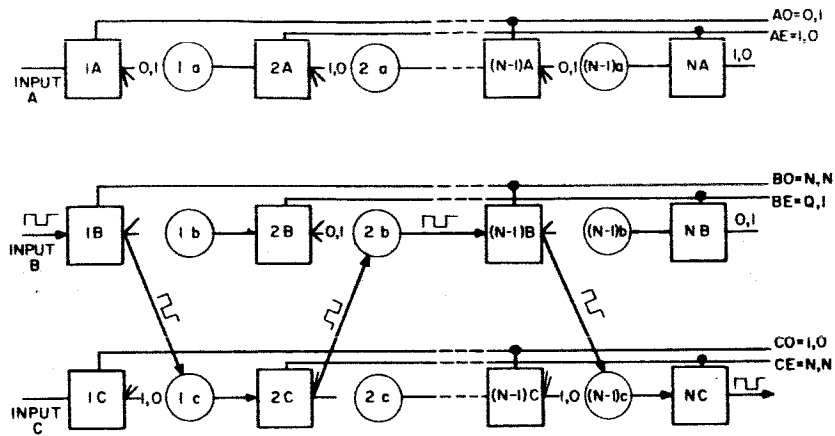


Figure 18a

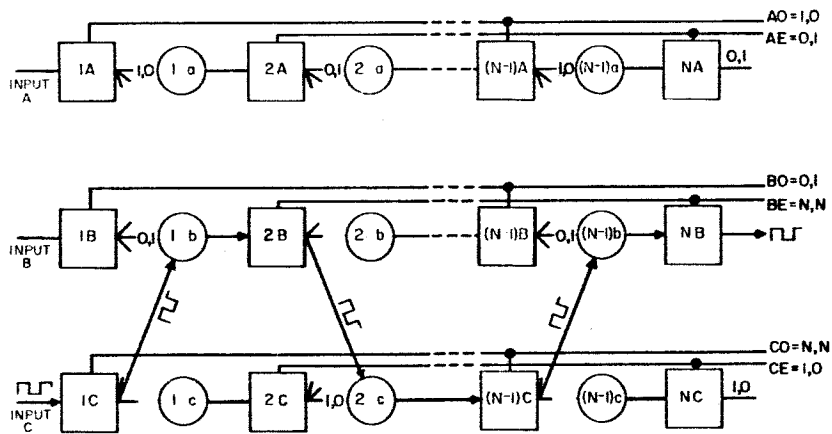


Figure 18b

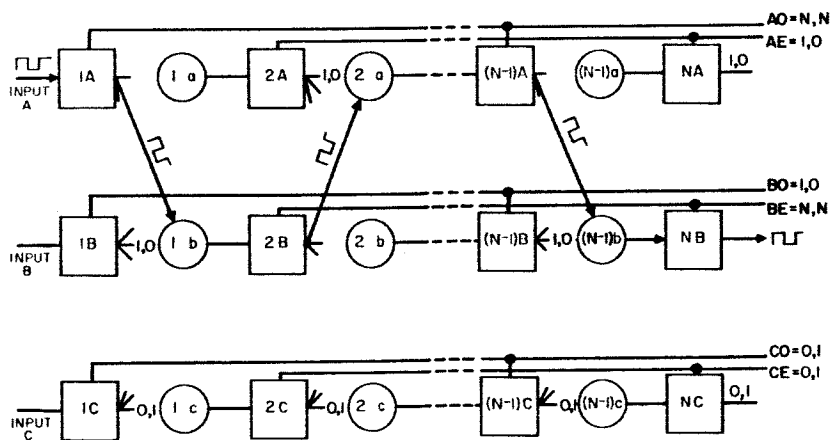


Figure 18c

Figure 18 Interwoven Rank Testing

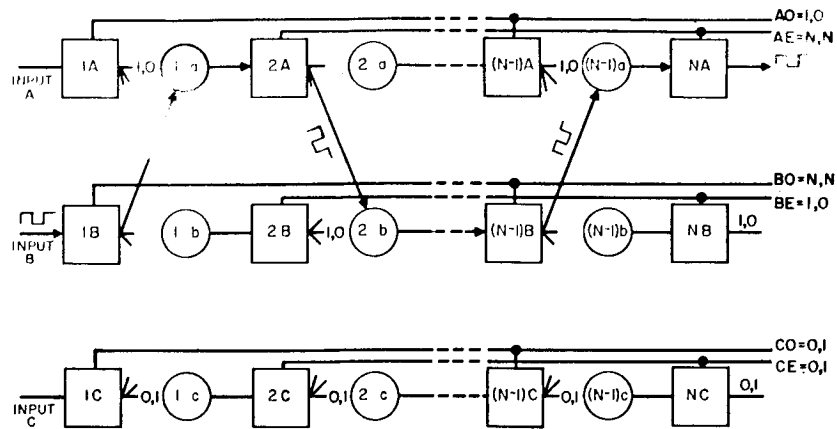


Figure 19a

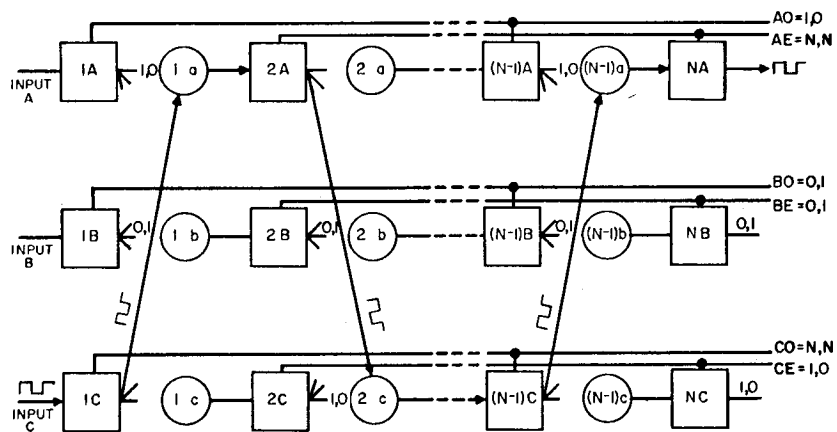


Figure 19b

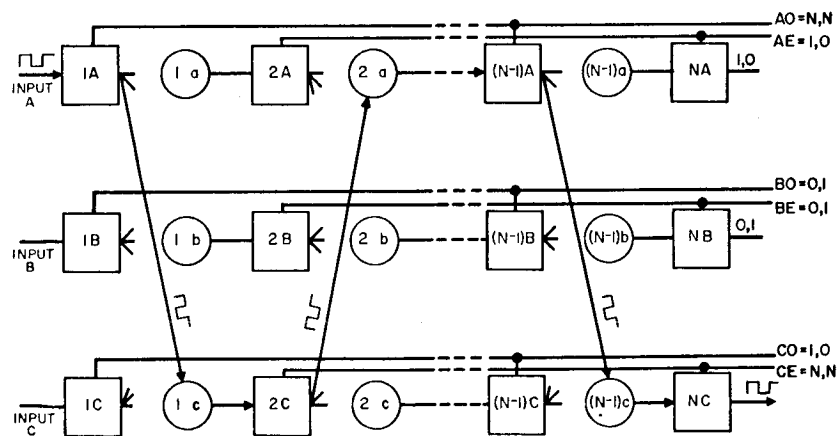


Figure 19c

Figure 19 Interwoven Rank Testing

processors and voters in the path operate correctly the final output of the Nth processor (NC) will be the correct output signal. Reversing the states of control lines AO, AE, BE, CO should also provide the same result since this causes the pairs of signal processors in each file to assume the opposite complementary condition. The system may be completely exercised as a non-redundant system for either of the above DC states.

Consider now the various combinations of input signals which the 1c voter was subjected to as a result of the above tests. An examination of figure 18a reveals that these combinations are as follows:

<u>State No.</u>	<u>A</u>	<u>B</u>	<u>C</u>	<u>Output</u>
3)	0	1	1	1
8)	0	0	1	0
7)	1	1	0	1
2)	1	0	0	0

Note that the tests have verified that the voter operated correctly for the two signal states which could not be confirmed by the basic singular rank tests. This was the uncertain condition that a voter will make a correct decision when the signal processor proceeding it in the same rank is in disagreement with the other two signal processors. Thus far our tests have verified the above uncertain condition for all odd numbered c rank voters, as well as all even numbered b rank voters. A total of four different input states have been verified for each of these voters. The remaining voters in these ranks may be similarly verified by the test conditions shown in

figure 18b. The a rank voters are verified by the arrangement shown in figure 18c and figure 19a. This is seen to be a mirror image extension of B-C rank tests.

At this point in the tests, the correct operation of all signal processors has been verified. An examination of the various input signal combinations which the voters were subject to is tabulated as follows:

<u>Rank a voters</u>			<u>Rank b voters</u>			<u>Rank c voters</u>		
<u>A</u>	<u>B</u>	<u>C</u>	<u>A</u>	<u>B</u>	<u>C</u>	<u>A</u>	<u>B</u>	<u>C</u>
0	1	1	0	1	1	0	1	1
0	0	1	0	0	1	0	0	1
1	1	0	1	1	0	1	1	0
1	0	0	1	0	0	1	0	0
			1	0	1			
			0	1	0			

Note that the b rank voters have been verified for six of the eight possible signal combinations while the a and c ranks were examined for only four.

Since the signal condition of all "1"s or all "0"s was previously shown to be trivial, it is evident that the b rank voters have been completely tested for proper operation under all combinations of input signals. The reason that only the b rank voters have been completely verified and not the a or c rank voters is due to the fact that the b rank voters provided a common signal path in the tests involving the c rank voters and the rank voters. The a and c rank voters may be completely verified by the tests shown in

: figures 19b and 19c. This is seen to cause the dynamic signal path to be interwoven between the a and c ranks.

Interwoven rank testing may therefore be used as an all inclusive procedure for detecting any failures of voters or signal processors without requiring access to any test points within the system. The system is reduced to sets of equivalent non-redundant systems by appropriate controls. It is then completely exercised and tested to determine if all functions are performed correctly. The success of all tests verifies that all signal processors and voters are failure free. If any of the tests result in an incorrect output, then some failure is present in the system. The detection of a failure gives very little information concerning its location within the system.

Although interwoven rank testing does not require access to test points within the system, it is a more elaborate approach which requires a degree of regularity in the system configuration as well as the establishment of twelve separate test conditions for an order three system, instead of the three required for singular rank testing and voter signal comparison. The system should be completely exercised for each of these tests to verify that the system is failure free if all tests are successful.

2. Failure Detection and Location for Maintenance

The alternate file controls described above may be used to detect and locate failures during normal system operation. Signal comparators are required only on the output of every signal processing file.

If a difference detector is integrally connected with each processor file, then the correct operation of the signal processors may be continuously monitored for maintenance purposes. If only test points are available, they may be periodically tested for signal disagreement. Any disagreement on the output of a signal processor will indicate that there is a failure in that signal processor or the voter which proceeds it. This failure may be repaired during system operation if the other replicated signal processor and voters in that file continue to operate correctly. If a module consists of one signal processor and the voter which provides its input, then repair is accomplished by replacing that module. This procedure is useful for detecting and locating failures which cause errors, but is not sufficient for determining the location of some failures within the voters. If all signal processors are failure free, the voter portion of the modules may be completely tested by imposing various combinations of signals at the voter inputs and examining the associated signal processor outputs for signal disagreement. To locate all possible voter failures, it is necessary to provide a means of examining signal processor outputs while subjecting the associated voters to the various combinations of input signals. This may be accomplished by controlling separately the odd and even files of the system or sub-system under test, as described in the previous paragraphs and illustrated in figure 18. For example, suppose that the odd files are allowed to operate normally and that each one of the three signal processors in the even files are in turn placed in each of the static DC states. The outputs of the odd files are monitored for signal disagreement during each

of the successive tests. Any disagreement on the output of an odd file signal processor will indicate that there is a failure in the voter which provides the input to that processor. Similarly, the outputs of the even files are monitored for each of the successive tests. Signal disagreement should be indicated whenever the control signal disagrees with the correct signal on the other processors in that file. If this indication does not occur, then either the control to that file is not effective, or there is a failure in the difference detector. The above testing is then repeated with the role of the odd and even files interchanged, each successive test examining the signal processors for disagreement. With proper design, any failures in the voters, the difference detectors, or the control hardware may be repaired while the system is in operation. Removal or disablement of one replicated voter or processor will not seriously jeopardize system reliability if the remaining replications of voters and processors continue to operate correctly.

D. Circuit Implementations

1. Control Circuitry

Consider now the mechanization for controlling the output of several signal processors with a single control line. A typical signal processor output is shown in figure 20. The circuitry shown is seen to be in the usual form of D-TL NAND gates. The base return resistor R_B may be connected to the emitter ground return if the associated transistor is representative of the low leakage silicon devices found in integrated circuitry. Since this resistor is normally connected to ground by a discrete

connective path, it is a relatively simple matter to provide R_B with a separate external connection.

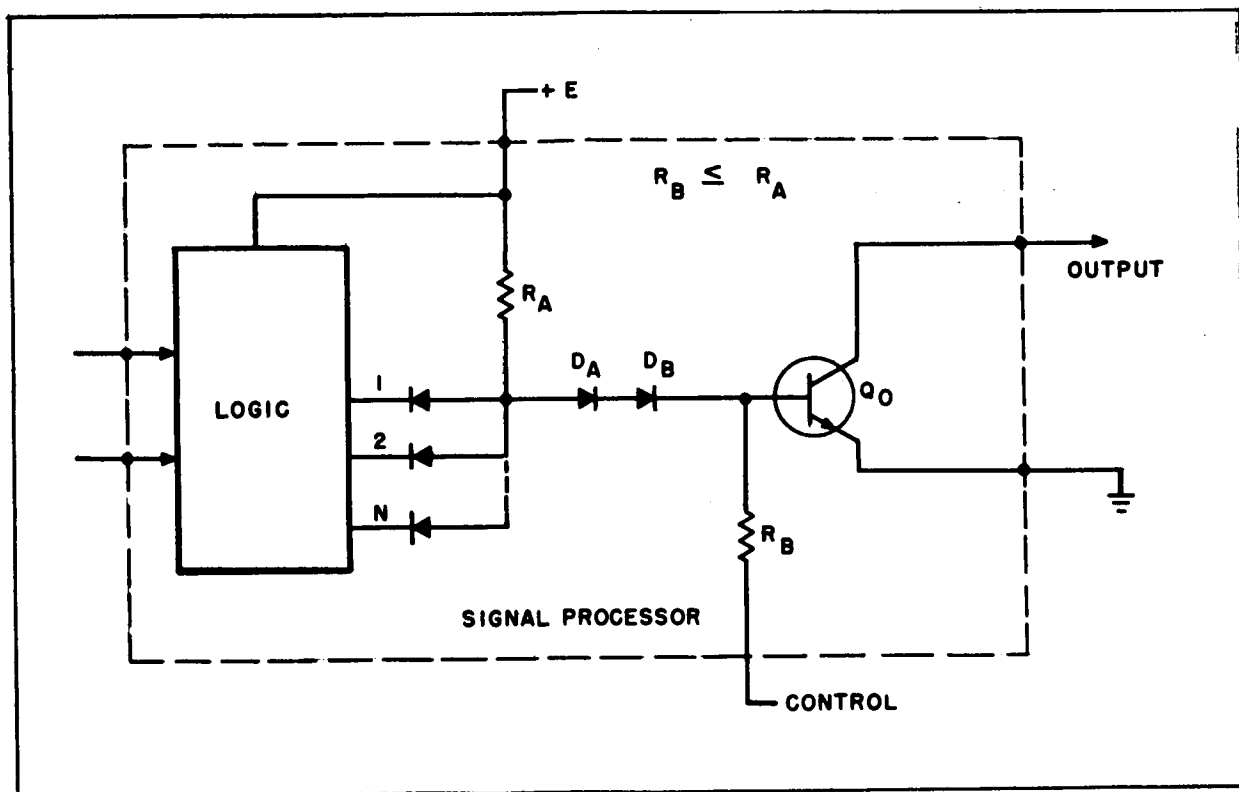


Figure 20 Signal Processor Output Control

Suppose further that R_B is chosen to be equal to or less than R_A . If R_B is connected to ground potential the circuitry will operate normally. If R_B is connected to the + E supply Q_0 will conduct and saturate regardless of the signals present on the inputs 1, 2, - - - N. This is seen to be the condition where the control line potential forces the signal processor output to assume the "0" state. If the control line is connected to an equal potential of opposite polarity (-E), transistor Q_0 will be cut off thus causing it to assume the "1" state regardless of the signals present on inputs 1, 2, - - - N. The method described to implement the required control function is one of several possible approaches. It is an approach which represents a simple modification to existing circuitry and requires only a single control line which is grounded in normal operation.

Another alternative requires control of both the base return line and the emitter ground line, but does not restrict the value of the base return resistor, R_B , and does not require a negative voltage supply. The same method described above is used to cause the "0" output, i.e., to connect the control line to a voltage which is sufficiently positive to cause the output to saturate. For most circuits, + E will be of sufficient magnitude for this purpose. To effect a "1" output, the emitter ground line may be removed, so that the output cannot be a low impedance to ground, regardless of input signals. This approach may be particularly useful when it would be undesirable to reduce R_B less than R_A , or in circuits where the base input diode, D_B , is replaced by an emitter follower to increase base current drive. This approach places little restriction on circuit

configuration or values and the test power supplies, but requires two separate control lines, both of which are grounded in normal operation.

2. Difference Detector Circuit

Shown in figure 21 is a typical discrete component difference detector which may be utilized in the foregoing tests. The output level is a logical "0" only if all inputs are identical. Any disagreement of input signals will cause the first transistor to conduct and thus cause the second transistor to assume the "1" state (cut off). The circuit is seen to perform the functional operation of "exclusive OR" for two inputs.

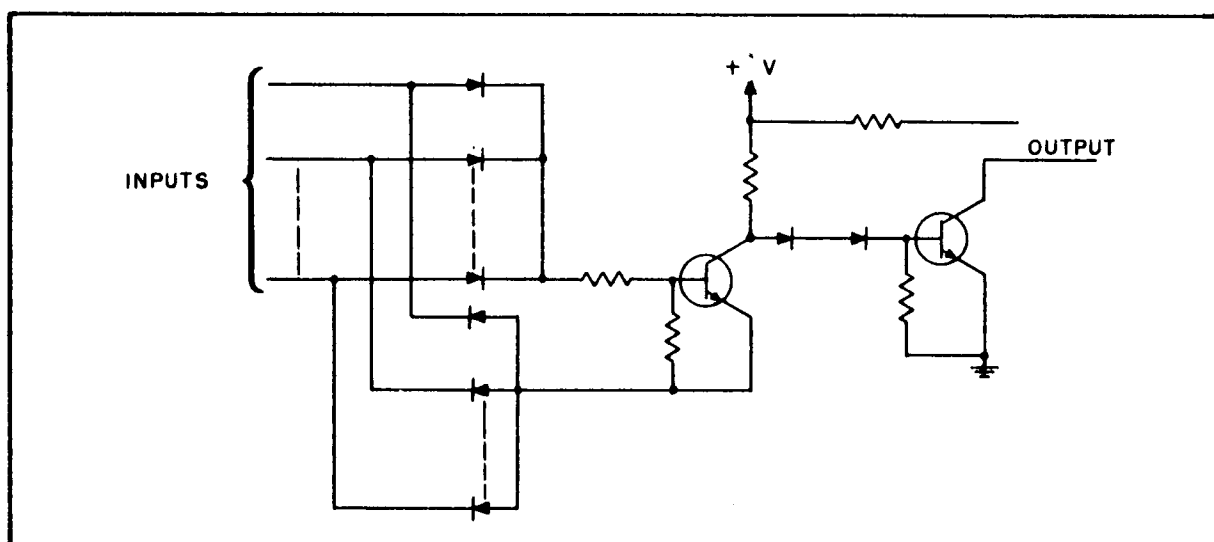


Figure 21 Difference Detector

The output of the difference detector may be used to trigger a flip-flop in order that any momentary disagreement of input signals may be displayed. This would be useful in detecting any sporadic errors which might otherwise remain unnoticed. As previously mentioned, the difference detectors might be combined with suitable indicators and packaged as an integral part of the system circuitry. This would eliminate any loading effects due to the use of test leads and external test equipment in monitoring test points. In addition this would provide maintenance personnel with a simultaneous display of the condition of the system and the location of faulty modules.

V. Summary and Conclusions

1. General

It has been shown that the special features of a redundant configuration impose unique requirements on the design of functional circuitry and the facilities required for test. Redundancy is a powerful tool for achieving extended reliability, but it should not be encumbered with circuitry which is inherently unreliable or contain particular failure modes which prevent the associated system configuration from operating independently. An appreciation of this philosophy allows the achievement of reliability goals with a minimum of additional complexity. Effective circuit design is required to obtain the desired balance between complexity and reliability in redundant systems.

2. Magnetic Logic

Although magnetic logic is often cited as having several features particularly applicable to spaceborne computers, the disadvantages of magnetic logic strictly limit their usefulness in general logic systems, and particularly for redundant spaceborne systems. Some basic disadvantages are listed below:

- 1) Lack of compatible steady output signals
- 2) Excessive power consumption for speeds comparable to low-power microcircuitry.
- 3) Extensive peripheral equipment, including high current drivers.
- 4) Limited fan-out and gain characteristics

5. High peak power requirements.
6. Indeterminate reliability performance due to extensive hand wiring with fine wire and numerous connections, as well as unavailability of accurate reliability data.
7. Complexity required for general logic functions.
8. Lack of suitable restoring element for use in redundant systems.

Magnetic logic does, however, offer non-volatile storage and reduced average power for low computing speeds. Magnetic devices appear to be suited to special applications where certain logic functions, such as transfer and OR, are intermixed with the memory function, and very low speed capability is acceptable.

3. Integrated Semiconductor Logic

Integrated semiconductor circuitry offers many characteristics which are desirable for circuits to be used in redundant space-borne systems. Some general features of integrated semiconductor logic when compared to other commonly available logic systems are:

1. Significantly reduced size, weight, and power consumption.
2. Availability of general logic elements, as well as special purpose circuits.
3. Predictable operating characteristics over wide environmental variations.
4. Availability of accurate reliability data.

5. Extensive research and development for new integrated circuits.
6. High frequency capability.
7. Compatibility with synthesis and testing techniques for redundant systems.

A comparison of the currently available integrated logic elements indicates that diode-transistor logic (D-TL) is the most suitable for use in redundant spaceborne systems. D-TL offers excellent operating characteristics, such as easily distinguished "1" and "0" states resulting in high DC stability and compatible output signals, high noise immunity, self contained drive current, allowable parameter tolerances, input isolation, and other characteristics which permit efficient redundant design. D-TL frequency capability exceeds the requirements of most spaceborne systems, and requires relatively low power, so that total power dissipation and temperature stress are minimized.

A majority voting restorer, designed using interconnected NAND elements, has been described which is not subject to the detrimental failures of conventional majority voters.

4. Failure Testing

It is a characteristic of redundant systems that they offer a

high reliability for a period of time after the initially failure free condition, and that the system reliability decreases rapidly when internal failures are present. It is therefore important to insure that no initial failures exist in a redundant system to obtain maximum system reliability. This reliability may be required for a single time interval without further maintenance, such as for spaceborne systems, or it may be required for repeated time intervals, where the system is restored to the initially perfect condition prior to each interval. The latter method may be used to obtain high mission reliability by maintaining a redundant system which is used repetitively, such as the ground support and launch equipment used prior to and during each mission. Since an initially failure free order three system can withstand any single failure, as well as a relatively large number of randomly scattered failures, it offers high reliability for the period of time when the probability of individual failures is low. Techniques are described which permit even higher reliability by combining periodic maintenance with continuous maintenance of a redundant system.

It has been shown that a relatively simple test referred to as singular rank testing may be used to determine that all of the replicated signal processors are working properly. If the signal processor fails whenever any of its parts fail, success of the singular rank tests will verify that all signal processors are failure free. Success of singular rank testing will also verify that the majority voters are sufficiently failure free to insure that the system is not vulnerable to single failures. Singular rank testing effectively isolates each rank of the replicated non-

redundant system by forcing each remaining pair of replicated ranks to have static complementary binary outputs. System output is monitored to determine if each individual rank is able to perform all system functions correctly, in a manner similar to the verification of a non-redundant system. Singular rank testing is expected to be the most efficient and effective method for diagnosing equipment which has been recently assembled from completely tested modules, since the probability that the few undetectable failures might have occurred since complete testing is very low.

A somewhat more complicated testing procedure, referred to as interwoven rank testing, has been described which will completely test all voters to insure that they will make correct decisions for all possible input combinations. It has been shown that the failure detection procedures may be accomplished by controlling one or more normally grounded common lines for each of the replicated ranks of the system, without altering the logic design or including any additional hardware except to provide access to these lines. Singular rank testing places no restrictions on system size or configuration.

The characteristics of redundant systems have been shown to introduce unique properties to the problem of failure location and faulty module replacement. Although a redundant system is more complex than its conventional counterpart, failure location within an operating system does not require the operator skill and simulation equipment usually required to locate failures in a non-redundant system. Since an operating redundant system always has at least one correct signal available at every point in the system, these correct signals may be used as a basis of comparison to

: other versions of the nominally identical signal. A difference detector on the signal processor outputs to restorers may be used to indicate failures among these signal processors. If the detector includes memory, it will also detect and locate transient or sporadic failures. These same difference detectors may be used for the somewhat more difficult task of locating those failures in the voters which do not cause errors when all voter inputs are identical, as well as verification that the test controls are actually capable of proper operation. The method which has been described uses the same types of control as singular and interwoven rank testing, and does not jeopardize system operation if all signal processors are operating correctly.

BIBLIOGRAPHY

1. Haynes, J. L., "Logic Circuits Using Square-Loop Magnetic Devices: A Survey", IRE Trans. on Elec. Computers, Vol. EC-10, No. 2 (June 1961)
2. H. D. Crane, "A High Speed Logic System Using Magnetic Elements and Connecting Wire Only," Proc. IRE, Vol. 47, pp. 63-73; (Jan. 1959).
3. D. R. Bennion and H. D. Crane, "Design and Analysis of MAD Transfer Circuitry," Proc. 1959 Western Joint Computer Conf., San Francisco, Calif., pp. 21-36, (March 1959).
4. J. A. Rajchman, "The Transfluxor," Proc. IRE, Vol. 44, pp. 321-332; (March 1956).
5. H. D. Crane, "Design of an All-Magnetic Computing System," IRE Trans. on Elec. Computers, Vol. EC-10, No. 2 (June 1961).
6. "Aviation Week and Space Technology," Aug. 19, 1963 pp. 93-103
7. A. R. Helland and W. C. Mann, "Failure Effects in Redundant Systems" Westinghouse Report EE-3351. (March, 1963)
8. Report No. NADC-EL-6319, Micro-Notes No. 3, "Information on Micro Electronics for Navy Avionics Equipment" (June, 1963)

Appendix 2

RELIABILITY OF IMPERFECT REDUNDANT SYSTEMS

by

R. S. Bray

P. A. Jensen

C. G. Masters

September 1963

TABLE OF CONTENTS

I.	INTRODUCTION	2-1
II.	MISSION RELIABILITY	2-2
III.	PROCEDURES FOR ESTIMATING THE SYSTEM RELIABILITY	2-6
	A. Estimation of the Expected Value of Mission Reliability with only the Information that the System is Operating at t_1	2-6
	B. Estimation of the Expected Value of Mission Reliability with Tests at t_1 Helping to Establish the Circuit Failure Rates	2-8
	C. Improvement of the Estimate Through Failure State Tests	2-9
	D. Determining the Mission Reliability of Large Systems	2-12
	E. Using Tests to Determine Both the Failure States of the System and Failure Rates of the Circuits at t_1	2-16
IV.	TEST OF THE HYPOTHESIS THAT MISSION RELIABILITY IS GREATER THAN A REQUIRED VALUE	2-17
V.	CONCLUSIONS AND RECOMMENDATIONS	2-19

I. INTRODUCTION

The problem of the pre-launch testing of spaceborne electronic systems is becoming more severe as the systems increase in complexity while decreasing in physical size. The testing problem will soon become much worse as systems are made redundant and in-flight tests are used to determine the successive actions of deep space probes. Tests can no longer be made adequately on the basis of a strict "working" or "failed" criterion because a redundant system may contain many internal failures and still be operating at the time of test. Such a system might easily have a much lower probability of successfully completing a mission than a functionally identical non-redundant system.

In addition, the large number of subsystems in a complex redundant network will make complete check-out (i. e. tests of each subsystem) virtually impossible. Consequently, a new method must be devised which will permit a statistical estimate to be made of the probability of mission success (reliability). This estimate must be based on the results of a limited amount of testing and should be as accurate as possible.

II. MISSION RELIABILITY

The problem may be stated more specifically as follows. A test of a redundant machine will be made at some time t_1 . (It is expected that some failures will be found in the equipment, and the object of the test is merely to determine the number and pattern of the failures in the system.) From the test data, the probability that the redundant system under test will operate successfully throughout a mission which begins at time, t_1 , and ends at time, t_2 , given that the system is operating at t_1 , is estimated. This probability is defined as the mission reliability (R) and is a function of the system organization, the state of the system at t_1 , the failure rates of the parts of the system, the starting time (t_1) of the mission, and the mission's duration, $t_2 - t_1$. At some time t_0 , which is less than t_1 or t_2 , all circuits in the system are assumed perfect. As time progresses they are assumed to fail in a random manner with a constant failure rate. At t_1 when the system is ready to begin the mission, the system must be in one of a finite number of possible failure states. The failure states are determined by the number and location of failed circuits in the system. For example, consider the multiple-line redundant network of figure Q-1. A restoring circuit indicated by a circle will make a correct decision if at least two of its inputs are correct.

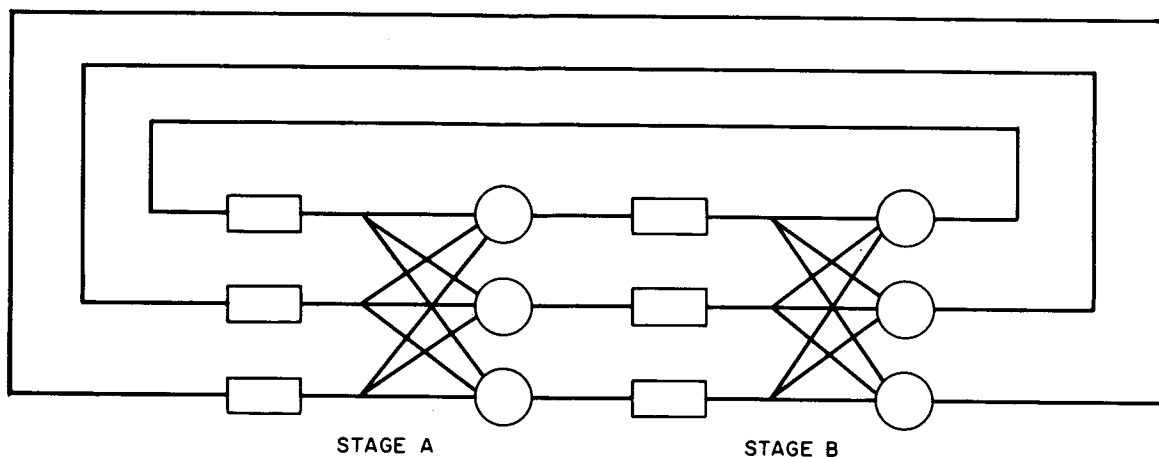


Figure Q-1. A Two Stage Example of a Redundant System

Assume for simplicity of explanation, that the restoring circuits of this system are perfectly reliable and that only signal processing circuits, indicated by rectangles, can fail. The possible failure states of this system are listed in columns 2 and 3 of Table I.

TABLE 1

1 Failure State	2 Number of Failures in Stage A	3 Number of Failures in Stage B	4 $R_i^* (t_2)^{**}$	5 $P_i (t_1)^{***}$
1	0	0	$\left[p_m^3 + 3 p_m^2 (1-p_m) \right]^2$	$\left[p^3 \right] \left[p^3 \right]$
2	0	1	$\left[p_m^3 + 3 p_m^2 (1-p_m) \right] p_m^2$	$\left[p^3 \right] \left[3 p^2 (1-p) \right]$
3	0	2	0	$\left[p^3 \right] \left[3 p (1-p)^2 \right]$
4	0	3	0	$\left[p^3 \right] \left[(1-p)^3 \right]$
5	1	0	$\left[p_m^3 + 3 p_m^2 (1-p_m) \right] p_m^2$	$\left[3 p^2 (1-p) \right] \left[p^3 \right]$
6	1	1	p_m^4	$\left[3 p^2 (1-p) \right] \left[3 p^2 (1-p) \right]$
7	1	2	0	$\left[3 p^2 (1-p) \right] \left[3 p^2 (1-p)^2 \right]$
8	1	3	0	$\left[3 p^2 (1-p) \right] \left[(1-p)^3 \right]$
9	2	0	0	$\left[3 p (1-p)^2 \right] \left[p^3 \right]$
10	2	1	0	$\left[3 p (1-p)^2 \right] \left[3 p^2 (1-p) \right]$
11	2	2	0	$\left[3 p (1-p)^2 \right] \left[3 p (1-p)^2 \right]$
12	2	3	0	$\left[3 p (1-p)^2 \right] \left[(1-p)^3 \right]$

* $R_i(t_2)$ is the probability of correct system operation at time (t_2) given the i^{th} failure state exists at t_1 .

** All the p_m 's in this column are probabilities that a circuit is successful at t_2 , given it was successful at t_1 .

*** All the p 's in this column are probabilities that a circuit is successful at t_1 , given it was successful at t_0 .

TABLE 1 (Cont)

1 Failure State	2 Number of Failures in Stage A	3 Number of Failures in Stage B	4 $R_i * (t_2)^{**}$	5 $P_i (t_1)^{***}$
13	3	0	0	$\left[(1-p)^3 \right] \left[p^3 \right]$
14	3	1	0	$\left[(1-p)^3 \right] \left[3p^2 (1-p) \right]$
15	3	2	0	$\left[(1-p)^3 \right] \left[3p (1-p)^2 \right]$
16	3	3	0	$\left[(1-p)^3 \right] \left[(1-p)^3 \right]$

* $R_i(t_2)$ is the probability of correct system operation at time (t_2) given the i th failure state exists at t_1 .

** All the p_m 's in this column are the probability that a circuit is successful at t_2 , given it was successful at t_1 .

*** All the p 's in this column are the probability that a circuit is successful at t_1 , given it was successful at t_0 .

For each of the failure states of Table 1, the reliability of the system can be calculated at t_2 . This is done as follows: If the failure rate, λ , of a circuit is constant and known, the probability that a circuit is successful at t_2 , given it is successful at t_1 is the exponential.

$$p_m = e^{-\lambda(t_2 - t_1)} \quad (1)$$

For the system to be successful at the end of the mission, two or three circuits in each stage must be successful. The probability that the system meets this requirement depends on the failure state of the system at t_1 , and the value of p_m . For instance for failure states 3, 4, 7, 8 and 9-16, the probability of correct system operation must be zero because there are too many failures at t_1 . Because R_i is defined as this probability, given the system is in the i th state at t_1 :

$$R_i = 0 \quad \text{for } i = 3, 4, 7, 8, 9-16$$

For failure state 1, the reliability is the probability that two or three circuits are successful at t_2 . Thus:

$$R_1 = \left[p_m^3 + 3 p_m^2 (1 - p_m) \right]^2$$

The reliability of the system for other failure states is shown in column 4 of Table 1.

Column 5 of Table 1 lists the probabilities that the particular failure states will be present at t_1 . The factor p in this column is the probability of success of a circuit at t_1 given the circuit was successful at t_0 . These probabilities will find use in later discussions.

Two things must be known if the mission reliability of the system is to be determined with 100% confidence, the failure state of the system and the failure rates of the circuits (needed to calculate p_m). For large systems both these factors may be very difficult or impossible to determine exactly. To find the failure state of a system, the failure state of each stage must be known. This may require a considerable amount of testing, probably a test of all circuits in the system. The failure rates of the circuits can only be determined exactly with a test of an infinite number of circuits all operating under the same environments as the circuits in the system. Of course, with limited testing allowed at t_1 it is improbable that the exact failure state of the system can be found. Estimates and their accuracy are the subject of the remainder of this report.

III. PROCEDURES FOR ESTIMATING THE SYSTEM RELIABILITY

In the study of this problem, several ways have been proposed to estimate a system's mission reliability with varying degrees of accuracy and varying levels of confidence. Four of these are described below.

A. ESTIMATION OF THE EXPECTED VALUE OF MISSION RELIABILITY WITH ONLY THE INFORMATION THAT THE SYSTEM IS OPERATING AT t_1 .

Using the design failure rates* one can estimate the mission reliability with only the information that the system is operating successfully at t_1 . This is done using the equations representing the reliability of the system at time t given only that all circuits are operating successfully at time 0. The system reliability $R(t)$ can be written as the probability of successful operation from time 0 to time t . The reliability of the system of figure 1 is:

$$R(t) = \left\{ p(t)^3 + 3 \left[p(t) \right]^2 \left[1 - p(t) \right] \right\}^2 \quad (2)$$

$$\text{where } p(t) = e^{-\lambda t}$$

A plot of $R(t)$ for the redundant system of figure Q-1 is shown in figure Q-2a.

* The design failure rates are those assigned to the circuits during the design of the system. They are generally derived from controlled life testing of components similar to those used in the circuits or from field tests of similar components.

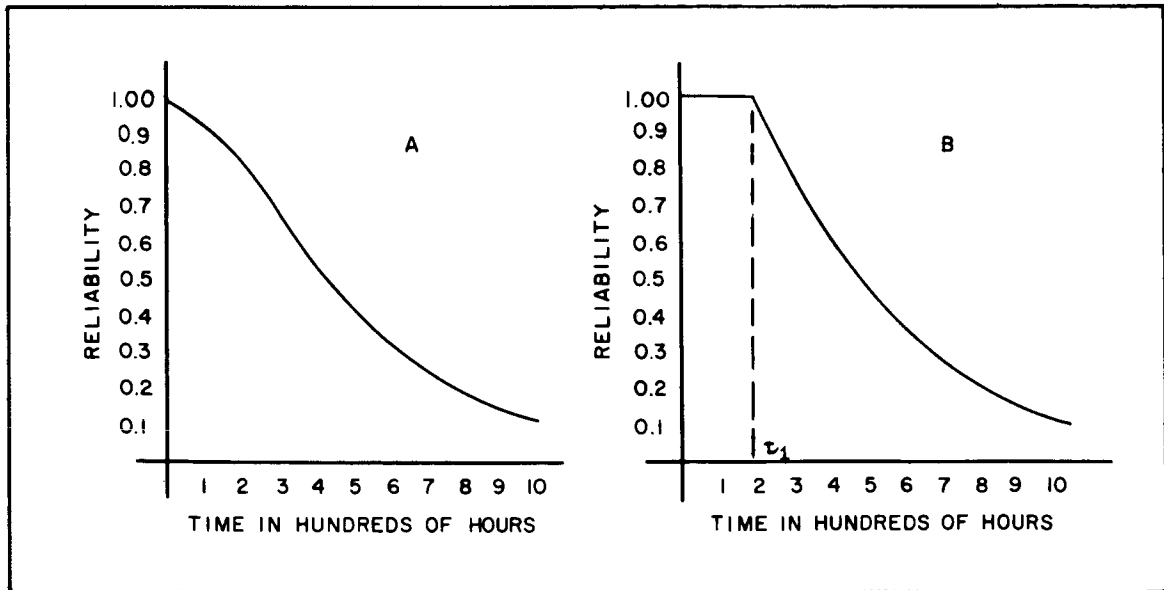


Figure Q-2. Reliability vs Time For a Redundant System.

A) With No Test at t_1 .

B) With a Test Determining the Success of the System at t_1

If one tests the system at a time t_1 and finds it to be working successfully, this information can be used to adjust the system reliability for time greater than t_1 to take account of the condition of success at t_1 . A curve must now be determined which gives the reliability of the system given successful operation at t_1 . This is expressed as:

$$R \left[t \mid R(t_1) \right]$$

For $t < t_1$, the reliability must be unity, because it is assumed that once a system fails it stays failed.

Then:

$$R \left[t \mid R(t_1) \right] = 1 \quad t < t_1 \quad (3)$$

For $t > t_1$, the reliability is:

$$R \left[t \mid R(t_1) \right] = \frac{R(t)}{R(t_1)} \quad t > t_1 \quad (4)$$

This is derived from the definition of conditional probabilities.

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

A plot of equations (3) and (4) is shown in figure Q-2b for a particular t_1 and the system shown in figure Q-1.

Using equation (4) the mission reliability can be written:

$$R(t_2, t_1) = \frac{R(t_2)}{R(t_1)} \quad (5)$$

Thus, the mission reliability can be determined simply by using the reliability equations of the system and the design failure rates of the circuits of the system.

The question now arises, of what value is this result? First, assuming the failure rates used in the calculation of R are perfect, if a large number of systems were constructed and run until t_1 , approximately $R(t_1) \times 100\%$ of them would be working. Throwing away all systems that were failed at t_1 and continuing the test until t_2 , $R(t_2, t_1) \times 100\%$ of the population all systems working at t_1 will be working at t_2 .

No information was given for this estimate about the failure state of the system at t_1 , except that the system was in one of the failure states for which the system is successful. For the example, these are states 1, 2, 5 and 6. This limited information about the failure state makes it necessary to approximate the mission reliability by an expected value given that the system is in one of the four successful failure states. The approximation has a considerable effect on the accuracy of the estimate which is described in detail in Section IIIC of this report.

B. ESTIMATION OF THE EXPECTED VALUE OF MISSION RELIABILITY WITH TESTS AT t_1 , HELPING TO ESTABLISH THE CIRCUIT FAILURE RATES.

Another problem which threatens the validity of the R calculated by this method is the uncertainty of the failure rates of the components of the system. The failure rates used in design are derived from a variety of sources and are almost surely not exactly accurate for any operational system. A realistic way to use design failure rates is to assign confidence limits to their values. With these one can say with a certain confidence that the failure rates of his parts are within a region determined by his confidence limits. This data is often available with design failure rates. Using the two extremes of failure rates, upper and lower confidence limits can be calculated for the mission reliability. The statement can then be made with a certain confidence that the mission reliability is within the interval of its confidence limits. It is instructive to point out that if the failure rates of all parts are perfectly known, there is 100% confidence in the calculated value of mission reliability. If, however, the failure rates are uncertain, as is always the case, confidence limits should be indicated for the mission reliability which reflect the uncertainty of the failure rates.

Estimation of the mission reliability of the system using the failure rates used in design has one serious failing. These failure rates often do not accurately describe the actual components. The design failure rates may have been determined under different environmental conditions than those of system in use, or components in the system may have been subjected to different manufacturing conditions than those used to derive the design failure rates. These and other factors might cause the circuits in the system to have different failure rates than those predicted in original design. Tests performed at t_1 can be used to determine if the actual failure rates are indeed different from design failure rates. If they are different the tests will be used to estimate the actual failure rate.

The first task is to test the null hypothesis that the actual average failure rates are the same as those used in design. To do this, the system must be split into groups of circuits with each group comprised of circuits of identical design. Using the design failure rates, the number of failures that can be expected in each group at t_1 is calculated.

This expected number is $p_j n$, where $p_j = e^{-\lambda_j^* t_1}$, and n is the number of circuits in the group. About this expected value one can construct an interval specifying the number of failures he is willing to observe at t_1 and still accept the hypothesis that the actual failure rate is that used in design.

The next step in the procedure is to test the circuits. If possible, all circuits are tested** and the numbers of failures recorded. If the number of failures at t_1 in n samples is within this interval the design failure rate is used to calculate the mission reliability. If the number of failures is not within the interval a new failure rate is calculated using the observed data at t_1 . The mean of this new failure rate is λ_o and is determined from the equation

$$\lambda_o = - \frac{\ln x/n}{t_1}$$

Confidence limits are placed on this calculated rate and the extremes of the confidence interval are used to calculate confidence limits on the estimates of the mission reliability of the system.

The question immediately arises, "Why test the null hypothesis at all if test data is to be accepted in preference to the design failure rates?" This is done because under the condition that the null hypothesis is met, the correspondence of the two sources of failure rate estimates would result in a higher confidence in the final estimate than either source alone can provide. When the null is rejected and the test data alone is used, the confidence in the estimate is reduced.

C. IMPROVEMENT OF THE ESTIMATE THROUGH FAILURE STATE TESTS

In this reliability estimation procedure a more accurate estimate is obtained by testing at t_1 to determine the failure state of the system. If the failure state were known exactly and the failure rates of the circuits were accurate, the mission reliability of the system could be calculated with no equivocation. Thorough testing at t_1 could determine exactly the failure state of the system, but since thorough testing is not of interest in this study the failure state will be known imperfectly. One will have a number of alternatives each with a certain probability given the results of the tests.

* λ_j = design failure rate of the j^{th} type circuit.

** Note, if the system is too large to permit complete testing, a random sample of each type of circuit is taken and the number of failures observed in the sample is used to estimate the actual failure rates.

Consider again the example of figure Q-1. Each stage of the system has four failure states, zero, one, two, or three failed circuits. If no information is available at t_1 , not even that the system is operating, every stage may be in any one of these states. Thus there are 4^2 possible failure states of the system. They have been listed in column 1 of Table 1. Associated with the i th failure state is a probability P_i which is the probability that the system is in this state at t_1 given that all circuits were successful at t_0 . Thus, with no information at t_1 on the condition of the system, the probability that the system is in the state in which no circuits have failed is

$$P_1 = p^6$$

The factor p is the probability of success of a circuit at t_1 . The probability of the failure state in which one circuit is failed in Stage B is

$$P_2 = 3p^5(1-p).$$

The probabilities of occurrence of the states given no information on the condition of the system at t_1 are listed in column 5 of Table Q-1.

Associated with each of the failure states is a reliability of the system at t_2 given that the system is in the failure state at t_1 . This is written as $R_i(t_2)$ and is shown for each state in column 4 of Table 1.

The reliability of the system is written as the sum over all i of the product of the probability of a i th failure state and the mission reliability given that the system is in the i th state at t_1 . Thus:

$$R(t_2) = \sum_{\text{all } i} P_i R_i$$

If tests are made at t_1 that give some information on the condition of the system, the number of failure states possible are markedly reduced, and the reliability estimate available at t_1 is much more accurate. For instance if one tests the system of figure Q-1 and finds it functioning correctly at t_1 , each stage must have no more than one circuit failure. Thus, only four states are possible after this test. These are states 1, 2, 5 and 6. The probability that the system is in a particular state must be adjusted to account for the known condition that the system functions at t_1 . Thus, for the example the probability of being in state 1 with no failures is:

$$\frac{P_1}{\sum_{i=1,2,5,6} P_i} \quad (6)$$

The denominator in equation (6) is the probability that the system is in one of the four possible states.

In general, a test to establish the failure state will leave only a set of possible failure states. Assume the test determines the state of the system to such an extent that the only possible failure states are included in the set I. If P'_i is the probability of being in the i th failure state given the results of the tests, then:

$$P'_i = 0 \quad \text{For } i \notin I$$

Or if a state is not in the set I its probability is zero.

If a state is possible then:

$$P'_i = \frac{P_i}{\sum_{\text{all } i \in I} P_i} \quad \text{For } i \in I \quad (7)$$

The mission reliability for a particular failure state, R_i , does not change, hence the mission reliability given the results of the test can be written in general as:

$$R_M = \sum_{\text{all } i \in I} \left[\frac{P_i}{\sum_{\text{all } i \in I} P_i} \right] R_i \quad (8)$$

For the example

$$R_M = \frac{1}{P_1 + P_2 + P_5 + P_6} \left[P_1 R_1 + P_2 R_2 + P_5 R_5 + P_6 R_6 \right] \quad (9)$$

More extensive tests at t_1 will further reduce the number of failure states which can exist. For instance if a test reveals that at least one circuit in the network is failed, the failure state which has no errors is eliminated, changing considerably the expected mission reliability. For this example $P'_1 = 0$, and states 2, 5 and 6 are the only members of the set I.

To illustrate the value of testing to determine the failure state at t_1 , consider the example. The probability that a circuit operates until t_1 is $p(t_1) = 0.9$ and the probability it lasts until t_2 , given it was successful at t_1 is $p_m(t_2) = 0.9$. The system is that shown in figure Q-1 and the restoring circuits are assumed perfectly reliable. Say that in reality one circuit is failed in one stage and the circuits in the other stage are all successful, but

this information is unknown to the tester. This is the information to be gained at t_1 through the tests. Table 2 lists the reliability one would predict with different amounts of information about the condition of the system at t_1 . The wide variation in the result indicates the importance of testing at t_1 .

This section does not propose the detailed procedures for testing a system at t_1 . It should, however, indicate the importance of making these tests and the calculations required to utilize the information gained from the test to estimate the system reliability.

TABLE 2

	Test Results at the Mission's Start (t_1)	Predicted System Mission Reliability	Corresponding Risk of Failure
1.	No information at t_1 , not even that the system is working.	0.821	0.179
2.	Tests show that the system is working at t_1 .	0.867	0.133
3.	Tests show that the system is working but that at least one circuit is failed.	0.770	0.230
4.	Tests show that exactly one circuit in the system is failed at t_1 .	0.788	0.212

D. DETERMINING THE MISSION RELIABILITY OF LARGE SYSTEMS

The example of the last section is a small two stage system. One might well ask if it is feasible to enumerate all of the possible failure states of a large system for the determination of the mission reliability. Indeed with no information at t_1 on whether or not an n stage system is operating correctly, there are 4^n possible failure states of the system. As n increases, the number of possible failure states increases exponentially.

The purpose of the tests at t_1 is to eliminate large numbers of these states in the manner shown for the example and hence obtain a better estimate of the mission reliability. The use of equation (8) provides this estimate but it requires, in its present form, separate consideration of each failure state. This is impractical for all but the smallest systems.

This problem is circumvented by first putting the mission reliability equation in a more general form. The mission reliability of the system given the results of the test at t_1 is a conditional probability which can be written:

$$R_M = \frac{\text{Prob. (Test results at } t_1 \text{ and successful system operation at } t_2)}{\text{Prob. (Test results at } t_1)} \quad (10)$$

Equation 8 is a representation of this equation for small systems.

The form equation (10) takes depends on the characteristics of the system under study and the type of test to which it is subject at t_1 . For example, consider an n stage order-three-multiple-line system which has perfect voters. For simplicity assume all the stages are identical with equally reliable circuits. For illustrative purposes assume the stages are arranged in a chain as in figure Q-3.

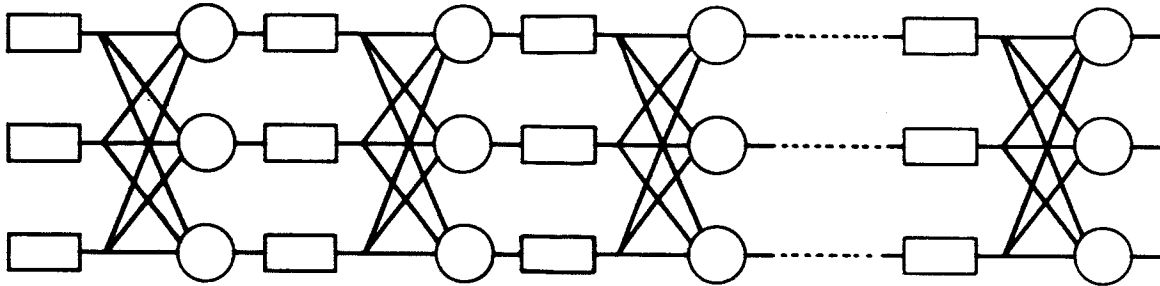


Figure Q-3. Chain of n -Multiple-Line Stages

The first type of test to which the system of figure Q-3 is subjected is a simple test to determine its operability. Is the system failed or successful at t_1 ? Given the system is successful at t_1 the mission reliability will now be determined.

Because the system is working at t_1 , each stage must be in one of two states, either three circuits successful or two circuits successful and one failed. Then the system may be in any one of 2^n possible states. Using equation (8) to evaluate the mission reliability would be a rather tedious and time consuming process if n were a sufficiently large value since both the numerator and denominator of this equation have 2^n terms. However, because of the independence of the stages of the multiple line system, it isn't necessary to carry out this operation. The probability that each stage is successful at t_1 is independent of the condition of all other stages and can be written:

$$\left[p^3 + 3 p^2 (1 - p) \right] \quad (11)$$

Since they are all identical the probability that all the stages are successful at t_1 is:

$$\left[p^3 + 3 p^2 (1 - p) \right]^n \quad (12)$$

This term is the probability that the system is in a successful failure state at t_1 and is the denominator for equation (10) when the test consists only of determining the operability of the system.

The probability that a single stage is operating at t_2 can be written:

$$\left\{ p^3 \left[p_m^3 + 3 p_m^2 (1 - p_m) \right] + 3 p^2 (1 - p) \left[p_m^2 \right] \right\} \quad (13)$$

Since the stages are independent the probability that system is operating at t_2 is:

$$\left\{ p^3 \left[p_m^3 + 3 p_m^2 (1 - p_m) \right] + 3 p^2 (1 - p) \left[p_m^2 \right] \right\}^n \quad (14)$$

This term is equivalent to the numerator of equation (10). Using the terms (12) and (14) the mission reliability can be determined for this system. Given that the system is successful at t_1 the probability that the system is successful at t_2 is:

$$R_M = \frac{\left\{ p^3 \left[p_m^3 + 3 p_m^2 (1 - p_m) \right] + 3 p^2 (1 - p) \left[p_m^2 \right] \right\}^n}{\left[p^3 + 3 p^2 (1 - p) \right]^n} \quad (15)$$

Note that for this determination of the mission reliability the separate failure states have not been enumerated. The calculation of mission reliability for this system has been a relatively simple procedure.

Other tests at t_1 will result in different forms for the mission reliability equation (10). For instance assume the system of figure 3 is subjected to a different test. This test subdivides the system into three nonredundant ranks as shown in figure Q-4.

Each rank will be tested individually. If a rank fails it can be inferred that one or more circuits in the rank are failed. If a rank is successful it can be inferred that all circuits in the rank are successful.

At t_1 the information is given that the system is operating correctly and that 0, 1, 2 or 3 of the ranks have failed. Now equations must be developed that determine the mission reliability of the system given the results of the test at t_1 .

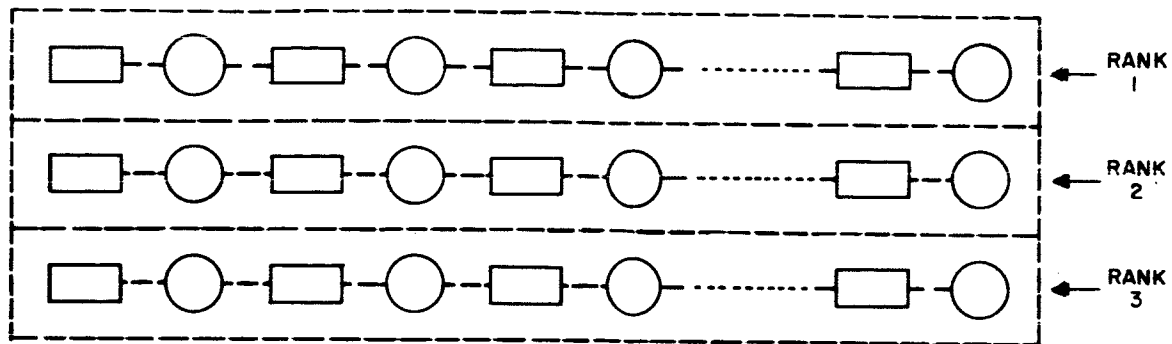


Figure Q-4. System Divided Into Three Nonredundant Ranks

The numerators and denominators of the mission reliability equation for the various test results are shown in Table 3.

TABLE 3

Test Result (Ranks Failed)	Prob. (Test Result at t_1)	Prob. (Test Result at t_1 and Successful System Operation at t_2)	Mission Reliability
0	$Y_0 = [p^3]^n$	$Q_0 = [p^3(p_m^3 + 3p_m^2(1-p_m))]^n$	$\frac{Q_0}{Y_0}$
1	$Y_1 = [p^2(1-p) + p^3]^n - Y_0$	$Q_1 = [p^2(1-p)p_m^2 + p^3(p_m^2 + 3p_m^2(1-p_m))]^n - Q_0$	$\frac{Q_1}{Y_1}$
2	$Y_2 = [2p^2(1-p) + p^3]^n - Y_0 - 2Y_1$	$Q_2 = [2p^2(1-p)p_m^2 + p^3(p_m^3 + 3p_m^2(1-p_m))]^n - Q_0 - 2Q_1$	$\frac{Q_2}{Y_2}$
3	$Y_3 = [3p^2(1-p) + p^3]^n - Y_0 - 3Y_1 - 3Y_2$	$Q_3 = [3p^2(1-p)p_m^2 + p^3(p_m^3 + 3p_m^2(1-p_m))]^n - Q_1 - 3Q_1 - 3Q_2$	$\frac{Q_3}{Y_3}$

Compared to enumerating all the failed states possible with the particular results of a test, these equations are relatively simple. If the assumption that all circuits are equally reliable is removed, the equations for mission reliability are very similar to these except instead of raising a single term to the power n as in these equations, a product of n factors will be taken. This should be a simple matter on a computer.

If the restriction that the restoring circuits be perfectly reliable is removed, the mission reliability equation will not be changed significantly unless the stages are interconnected in such a manner that they are no longer independent. The techniques used to calculate system reliability in this section are invalid if the stages are not independent. Techniques have been developed to determine the reliability of such systems* and these must be used in determining the mission reliability.

The equation describing the mission reliability for a system will depend on both the tests performed at t_1 and the characteristics of the system. These factors will surely be known prior to the test, so equations can be developed to evaluate the mission reliability which take into account the possible failure states of the system without exhaustive enumeration.

E. USING TESTS TO DETERMINE BOTH THE FAILURE STATE OF THE SYSTEM AND FAILURE RATES OF THE CIRCUITS AT t_1

In technique C, tests were made at t_1 to determine the possible failure states of the system. In technique B tests were made to establish the actual failure rate of the circuits of the system. It should be possible to design tests which give information regarding both these parameters.

The tests will establish the failure rate of the system at t_1 and use these in carrying out the reliability calculations described for Technique C. It takes little imagination to see that in the course of tests to determine the failure rate a great deal will be learned about the failure state of the system. For instance as soon as one failure is found the possibility that the system is in the no circuit failure state is decreased to zero, probably decreasing the mission reliability appreciably.

The details of this technique have not been developed, but generally it proposes to use the tests of t_1 to indicate both these parameters and thereby increase markedly the accuracy of the mission reliability estimate.

* Jensen, P. A., W. C. Mann and M. R. Cosgrove, "The Synthesis of Redundant Multiple-Line Networks", First Annual Report Contract NONR 3842 (00), May 1, 1963.

IV. TEST OF THE HYPOTHESIS THAT THE MISSION RELIABILITY IS GREATER THAN A REQUIRED VALUE

This method is separated from the others because it does not explicitly estimate the reliability of a system. Instead it finds, through measurements at the beginning of the mission, the probability that the system will not meet a given mission reliability specification.

The user of the system must specify the minimum mission reliability. He must also specify the maximum chance he is willing to take that the system does not meet this goal when his tests indicate that it will. It is assumed that the system is not acceptable if the probability that it does not meet the reliability specification is above the given value, and is acceptable otherwise.

The first step in this procedure is to determine the failure rates that the circuits of the system must have to just meet the mission reliability goal. These failure rates are called the maximum failure rates, λ_m . For a system in which many circuits have the same failure rate this does not seem to be too imposing a problem. For example consider a system where all circuits have the same failure rate. If the starting time and duration of the mission are known, the mission reliability can be expressed only as a function of the failure rate, λ . Equation (5) can then be set equal to the required mission reliability and solved for the failure rate. A cut and try method may be required for the solution.

The maximum failure rate is a function of both the starting time, t_1 , and the duration, $t_2 - t_1$, of the mission. However, if the duration of the mission is known, it is possible to plot a curve of mission starting time against the maximum failure rate.

Once the maximum failure rate is known it only remains to determine if the actual failure rate of the circuits of the system is less than or equal to this value. This will be determined by testing n of the circuits at t_1 and counting the number of failed circuits. Call the number of failed circuits X_1 . With this data and by using the maximum failure rate, an upper bound on the probability that the true failure rate is greater than the maximum failure rate can be determined.

If the fact that a majority of the circuits in a stage must be operative at t_1 is neglected, the success of a circuit in the system may be considered a Bernoulli trial with probability of success, $e^{-\lambda t}$. The probability distribution of the total number of circuit failures in M circuits is then binomial. This distribution or the associated density function can be plotted for any number of samples. One such plot appears in figure Q-5.

The probability distribution of the number of failures at time t_1 can be plotted using the calculated maximum failure rate.

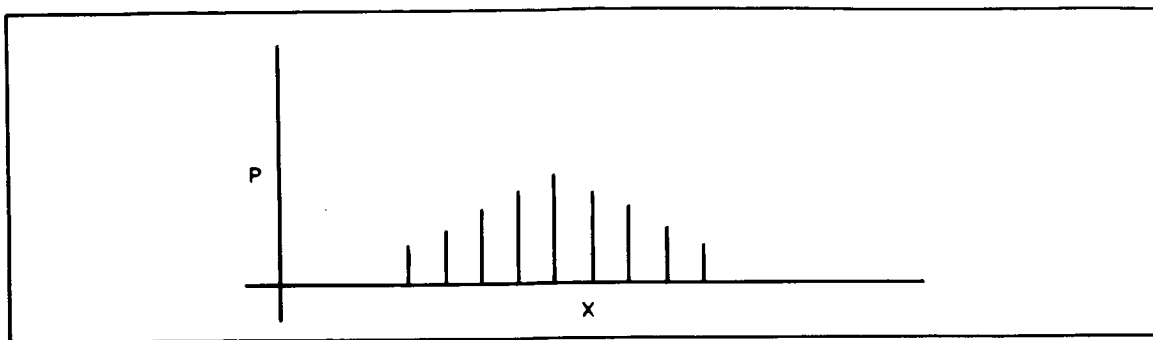


Figure Q-5. Sample Distribution

Some maximum number of failures Y will be chosen such that there is probability of δ that the number of failed circuits observed at t_1 , X_1 , will be less than Y if the failure rate of the circuits is λ_m . The quantity δ is determined from the binomial:

$$\delta = \sum_{h=0}^{Y-1} \binom{n}{h} (e^{-\lambda_m t_1})^{n-h} (1 - e^{-\lambda_m t_1})^h \quad (16)$$

For failure rates greater than λ_m the probability that less than Y failures occur must be less than δ . So if X_1 is less than Y , with confidence $1 - \delta$ the statement can be made that the actual failure rate must be less than the maximum failure rate. Now the statement can be made that with confidence $1 - \delta$ that the reliability of the system is greater than the minimum reliability specified by the user.

This method leads to the statement with a confidence $(1 - \delta)$, it can be said that the probability that the system will succeed is R . The information used to compute R might be used to compute the expected time to system failure instead. The object of the test would then be to confirm or reject the hypothesis that the expected life would exceed the mission time with a confidence $(1 - \delta)$. This modification has not been carefully examined but it appears to reduce the number of probabilistic statements from two to one.

This procedure again uses no information on the failure state of the system except that the system is successful at the beginning of the mission. The effect of this on the accuracy of the results has already been discussed in Section IIIC.

V. CONCLUSIONS AND RECOMMENDATIONS

It is the nature of a redundant system to withstand a number of internal failures and still perform its function successfully. This is an extremely desirable property for increasing life or providing high reliability, but it makes it unreasonable to base the decision — whether or not to carry out a mission with the system — only on the fact that the system is operating at the beginning of the mission.

This decision should be based on the probability that the system will complete the mission successfully. There are two major factors affecting the probability which are imperfectly known at the beginning of the mission. First, the number and location of initial circuit failures has a very significant effect on the probability that the system will operate throughout the mission. Second, the mission reliability depends heavily on the failure rates of the circuits which make up the system. There is little accurate information concerning either of these factors when it is time to make the decision.

The report proposes that certain tests be made just before the mission is to begin to determine at least approximately, these unknowns. It proposes some procedures for using the results of the tests to estimate the mission reliability with varying degrees of accuracy. A procedure for making the decision on the useability of the system without estimating the mission reliability is also presented.

It should be noted that the details of these procedures are still to be worked out and the accuracy of their results are still uncertain. The work here reported will provide the basis for future studies on the subject.

No attempt has been made to evaluate the relative usefulness of these procedures. It is recommended that efforts be made to develop an appropriate measure for comparing the techniques so that they may be evaluated relative to a common scale.

One very important area of study neglected by this report is the design of simple and efficient tests to be performed at the beginning of the mission to obtain the information required for the reliability estimates. As much information as possible must be gained from a minimum number of tests. A small amount of basic work has been done in this area, and it will be the subject of future efforts.

Appendix 3

A SURVEY OF COMPONENTS FOR ADAPTIVE RESTORING CIRCUITS

by

H. Brinker

TABLE OF CONTENTS

	Page
Introduction	1
1. Electrochemical Devices	3
a. Device 1	3
b. Solion	5
c. Mercury Cell	7
2. Magnetic Devices	8
a. MAD Integrator	8
b. Orthogonal Core Integrator	11
c. Second Harmonic Integrator	11
d. Magnetostrictive Integrator	12
3. Conclusion	13
References	15

LIST OF FIGURES

Figure 1	Comparison of Adaptive and Majority Voting Techniques	2
Figure 2	Adaptive Voter	2
Figure 3	Device 1 Cell	4
Figure 4	Device 1 Integrator	4
Figure 5a	Solion Tetrode and Output Characteristics	6
Figure 5b	Solion Tetrode connected as an Integrator	6
Figure 6	Mercury Cell Integrator (capacitive readout)	7
Figure 7	Multiple Aperture Device (MAD)	9
Figure 8	MAD Integrator	10
Figure 9	Orthogonal Core	11
Figure 10	Second Harmonic Integrator	12
Figure 11	Magnetostrictive Integrator	13

Introduction

The Adaline Neuron¹ is an adaptive logic device which may be trained to recognize certain classes of input patterns. The device output is a binary signal which classifies particular combinations of input signals into two categories. An output decision is determined by a threshold element whose input is the linear sum of the products of each input and its associated variable weight. During adaption the weights are appropriately changed in order to make the output decision agree with the desired response. By following a simple set of rules after each application of input signal combinations the device is caused to converge to an optimum state for properly categorizing the set of input patterns.

Although training rules for a single layer system have been formulated by Widrow^{1,2} new adaptive theory is required if systems of two or more cascaded layers are to be properly trained to perform complex functions of adaptive behavior and pattern recognition. The question of whether such devices may be connected in complex arrays and demonstrate brain-like behavior has generated considerable interest. Such applications appear to be philosophical and subject to considerable controversy. Of primary concern in the present study is to consider the usefulness of the Adaline neuron approach in implementing the adaptive voting elements of a redundant system.

The chart of Figure 1 shows how adaptive voters may extend the reliability of a conventional redundant system, allowing a system using 9 replicas to outperform a conventional system using 35 replicas of each function.

The Adaline neuron has received considerable quantitative study in application to pattern recognition. When modified as shown in Figure 2, and applied as an adaptive voter, the training rules become quite simple since the desired output is determined by a voting of the weighted inputs. Initially, all weights (gains) are made equal. The decision element will then provide an output in accordance with the states of the majority of binary, replicated input signals. If input errors are independent and random the adaptive voter, by progressively adjusting its weights to assign high weights to reliable inputs and low weights to failed or unreliable inputs, may derive correct information from a small minority of correct inputs.

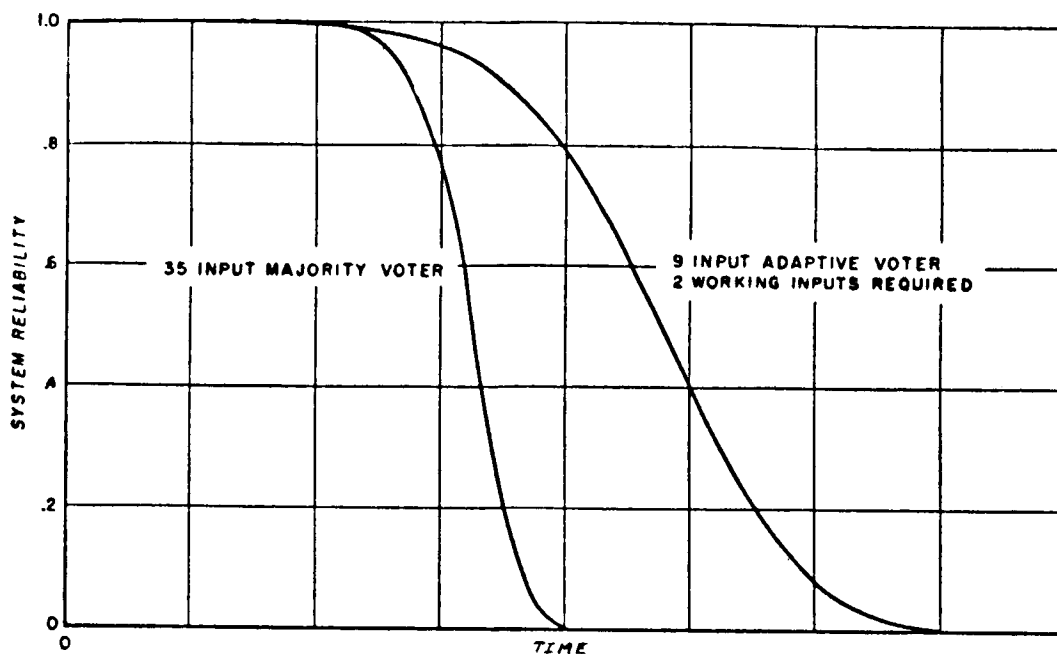


Figure 1 Comparison of Adaptive and Majority Voting Techniques

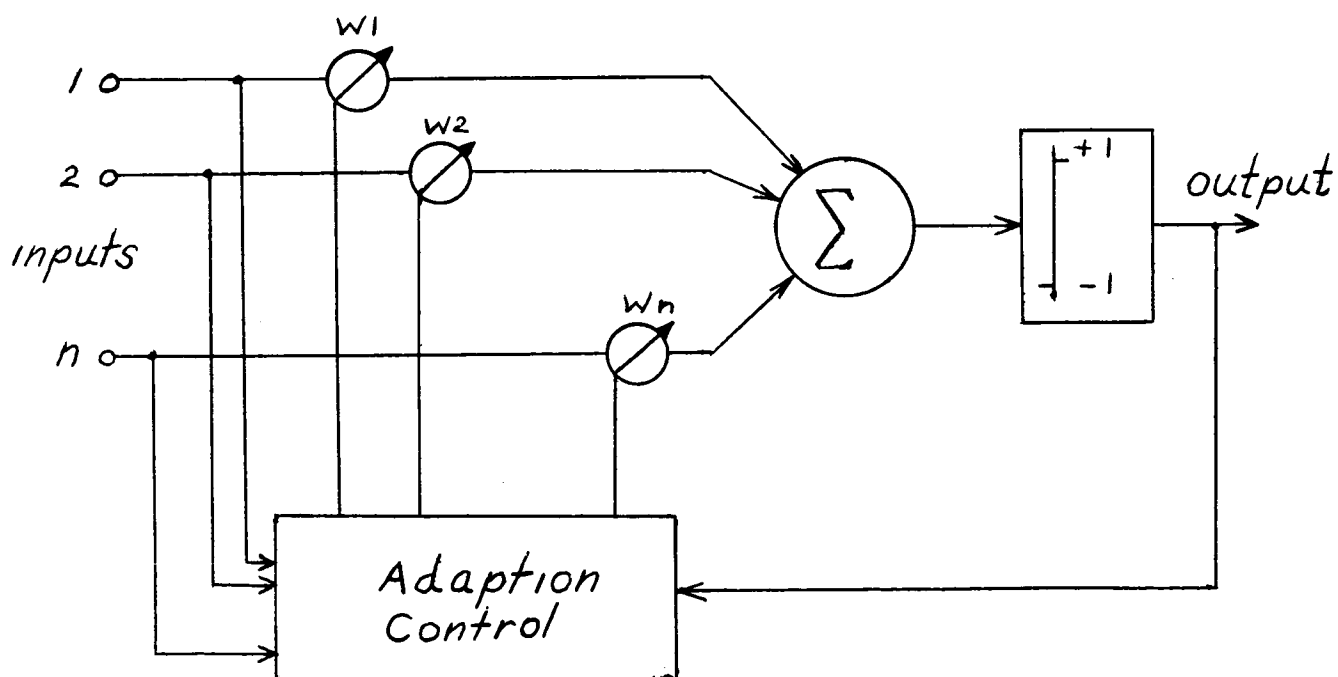


Figure 2 Adaptive Voter

In this manner the effect of errors caused by input failures may be negated, allowing a correct decision to be made under a high probability of input signal failure. The simple, fixed majority voter will make output decision errors when more than half of the inputs fail or are in error. The adaptive voter, by masking out input errors as they occur, may tolerate failures until only two correct inputs out of the original group are present.

In order to provide automatic adaption it is necessary to continuously compare the output decision with each binary input and to incrementally decrease or increase each input weight according to whether agreement or disagreement exists. Assuming that input errors or failures occur randomly and that the automatic adaptive process can negate an unreliable input before other failures occur, the adaptive voter offers the possibility of realizing system reliability of unprecedented excellence.

Inherent in the basic design of an adaptive voter is the requirement for a variable weighted device which performs integration and displays relatively permanent memory. These special characteristics have stimulated considerable effort toward the development of suitable adaptive components. Devices which display variable weight with memory generally utilize phenomena involving atomic translation or rotation. The following represents a survey of the more promising techniques which have been suggested by researchers. The first three devices described exploit electrochemical effects while the remaining devices utilize magnetic domain phenomena.

1. Electro-Chemical Devices

a. Device 1

Device 1³, an electrolytic device developed at Stanford University by Widrow, is an electronically adjustable resistor with a rate-of-change of resistance controlled by application of d-c current in a third electrode. It consists of a sealed plating cell containing an electrolytic bath, a resistive substrate upon which metal is deposited and a metal source electrode. A typical configuration indicating the placement of electrodes and electrolyte in a small plastic enclosure is shown in Figure 3. Two leads are attached to the substrate and resistance between these leads can be reversibly controlled by passing plating current into a third electrode. The conductance of the device is changed and stored by plating or stripping metal from the substrate by means of the integral of the plating current. Conductance is sensed nondestructively by applying a low voltage a-c signal and measuring the resultant current flow.

Normal d-c drop between source and substrate is typically 0.2 volts at a plating current of 0.2 ma. The substrate resistance changes from 30 ohms to 2 ohms in 10 seconds with this magnitude of plating current. The AC sensing voltage applied is usually 0.1 volts RMS. A typical implementation of Device 1 with associated transformer coupled sensing and d-c plating circuitry is shown in Figure 4.

Although Device 1 models are commercially available at a cost of approximately \$50 per cell their application in a practical system is somewhat cumbersome. Transformer coupled circuits are usually required in order to present a balanced load to the plating current source, and to provide the

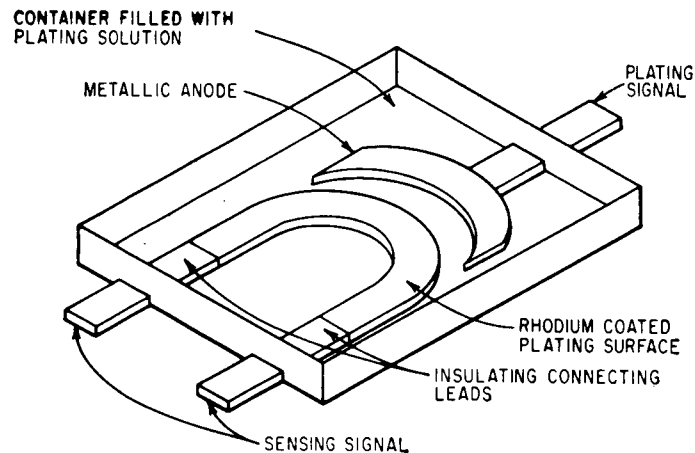


Figure 3 Device 1 Cell

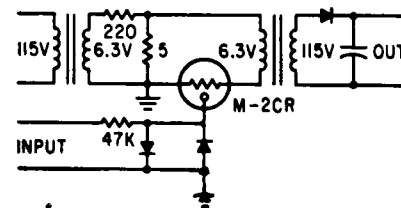


Figure 4 Device 1 Integrator

low voltage drop across the substrate. The substrate resistance is usually less than 100 ohms and the a-c voltage drop must be kept below $3/4$ volt in order to prevent the formation of gas in the cell. Some difficulty has been reported in keeping the substrate material free of dimensional imperfections which in turn cause non linear plating effects to take place. Long term stability is apparently affected by chemical reactions taking place between plating material and electrolyte. To date Device 1 models are available in sample quantities and it is difficult to predict ultimate large scale production costs, repeatability and reliability.

b. Solion

The solion is a fluid-state device which functions by controlling and monitoring a reversible electrochemical "redox" reaction. The term redox refers to a chemical reaction in which oxidation and reduction occur simultaneously. The redox system used in solions consists of two electrodes immersed in an electrolyte containing both the oxidized and reduced species of an ion. The system is completely reversible in that oxidation can occur at either electrode while an equivalent amount of the same element is reduced at the opposite electrode. Iodine is the reacting element most commonly used.

A simplified drawing of a solion tetrode and its output characteristics is shown in Figure 5a. The tetrode has a platinum electrode at each end of a glass tube and two perforated platinum electrodes separating the tube into three compartments. The reservoir, containing the input electrode, is the largest compartment. The integral compartment, containing the common electrode, is made very small so an equilibrium distribution of the iodine may be quickly reached. The compartment between the shield and readout electrodes serve to separate the two electrodes. The output characteristics of a solion tetrode are similar to that of a vacuum tube pentode, and show a transconductance of 40,000 micromhos at an output current of 500 microamperes.

A solion tetrode connected as an integrator is shown in Figure 5b. By controlling the charge transferred between the two input electrodes, a change in conductivity proportional to the integral of the input current may be obtained between the output electrodes. In this manner the device may be utilized as an integrator, providing an output current proportional to the integral of the input current. Because of the concentration potential, the input impedance of the solion tetrode is in the order of 1000 ohms and therefore a relatively high impedance signal source is required in order to avoid integration errors. At constant temperature, the stability of solions is reported to be less than 1% over a period of several days.

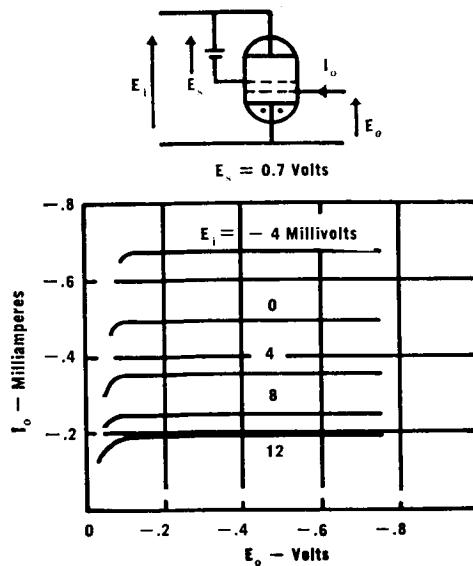


Figure 5a Solion Tetrode and Output Characteristics

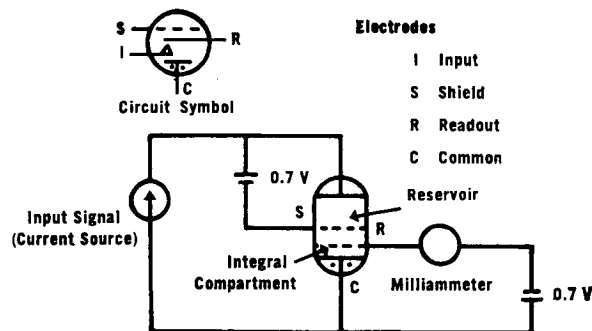


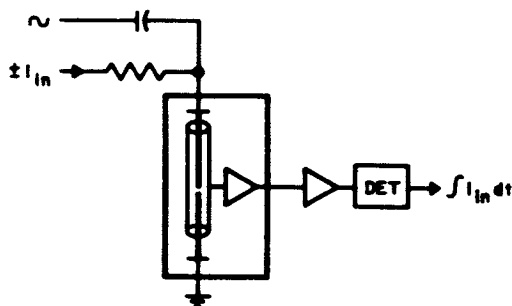
Figure 5b Solion Tetrode Connected as an Integrator

A practical problem in the use of solion tetrodes arises from the requirement of providing an isolated battery potential between input and shield electrodes to prevent iodine diffusion between the reservoir and integral compartments. Primary application for the solion tetrode to date has been demonstrated as a low level DC amplifier with a time constant of

20 seconds. Because of the inherent practical problems of precision design, isolated supply voltages and discharging effects of parallel outputs the solion appears to offer little promise as a practical adaptive component.

c. Mercury Cell

Another novel approach for variable gain with memory is achieved by use of a mercury cell integrator,⁶ an electrochemical device which provides visual and electrical readout of the integral of an applied current. The integrating element consists of a capillary tube filled with two columns (electrodes) of mercury separated by a gap of aqueous electrolyte of metallic salt. Two different methods have been used to provide electrical readout. The first method called capacitive readout is shown functionally in Figure 6. The d-c input signal electroplates mercury across the gap at a rate which is a direct function of the input signal amplitude, thus causing the gap or bubble of electrolyte to move. The outside of the capillary is covered by a vapor-deposited conductive sheath. The mercury electrodes and sheath, separated by a thin glass wall provide a capacitance of approximately 20 pF. In application, an a-c signal is connected across the electrodes and



CIRCUIT DIAGRAM

Figure 6 Mercury Cell Integrator
(Capacitive Readout)

superimposed on the d-c input signal. The a-c signal will divide in accordance with the capacitance existing between the upper mercury column and sheath, and the capacitance between sheath and lower grounded column of mercury. The excitation signal provides a signal at the sheath which is a direct function of the length of the ungrounded electrode. An auxiliary amplifier and detector in turn provide a proportional d-c signal of proper level to operate other related devices.

The device provides reversible integration, relatively stable memory, direct visual readout and a linearity better than 0.1 percent. Input control current is limited to +5 ma d-c. The integration time from minimum to maximum output signal is approximately 100 minutes at maximum control current. This time is ultimately limited by the maximum voltage which may be dropped across the electrolyte, without causing the formation of gas.

A typical capacitive readout integrator now commercially available is approximately 0.5 cu. in. but prices range around \$130 per unit. Although displaying excellent stability and predictable operation such devices will require considerable price reduction before application becomes practical. The integration time although relatively long may not present a serious limitation for systems which display slow adaptive behavior as would be the case in adaptive voting elements.

Another technique for sensing the position of the bubble utilizes a light source and a photo-conductor whose resistance is inversely proportional to the amount of light passed by the transparent electrolyte. As the bubble moves out of line with the light source and photo-conductor target area the light becomes progressively blocked by the mercury columns, causing the photo-conductor resistance to increase. This technique allows faster integration because the bubble need only be displaced by its own height to effect a change from maximum to minimum light intensity at the photo-conductor. A typical photoelectric integrator commercially available occupies 1 cu. inch and requires 300 milliwatts to power an integral incandescent lamp. Output resistance varies over the range from 25K ohms to 350K ohms. Quantity prices are expected to fall below \$15 per unit thus providing a reasonably inexpensive adaptive component. The use of an incandescent lamp for the light source imposes a serious life and reliability problem. The use of a more reliable light source and a substantial size reduction will be necessary before application becomes practical.

2. Magnetic Devices

Various techniques have been suggested for providing variable gain and non-destructive readout with magnetic devices. The phenomena utilized in such devices is based upon the ability of magnetic materials to store a remanent flux which is sensed in a non-destructive manner. Suggested devices provide the capability for a partial switching of magnetic domain under a volt-second impulse as the basic incrementing source. Suitable magnetic materials include ferrites and tape wound cores which are characterized by a square hysteresis curve. Most of the devices to be described utilize the same basic type of incrementing technique and differ primarily in the manner by which the stored flux is sensed.

a. MAD Integrator

A diagram of a typical multi-aperture device⁷ is shown in Figure 7. In this device flux can be switched around the minor aperture by means of an a-c drive winding without disturbing the flux linking and stored around the main aperture. Initially the flux around the main aperture is set to cause saturation in either a clockwise or counterclockwise direction. A momentary reversal of the magnetizing force driving the main aperture will cause a partial reversal of the flux. The amount of flux reversal is determined by the magnitude and duration of the drive and the value of the hold current. The purpose of the hold winding is to retain a portion of the core saturated in the original direction of magnetization and thereby assure partial switching of the flux. The amount of flux alternately switched around the small aperture is then proportional to the flux which has been switched

around the main aperture. The output voltage will consist of a signal whose voltage integral is proportional to the amount of flux trapped in the common area between the two flux paths. Several cycles of carrier drive may be required before this condition stabilizes. Care must be taken to limit the carrier drive to values less than the magnetizing force required to disturb the remanent flux around the main aperture.

The extent to which the remanent flux can be incremented is usually implemented by means of a smaller core of like magnetic material. The smaller core provides the appropriate amount of volt-second drive to increment the storage core in equal steps at various settings of remanent flux. Brain⁸ has indicated that it is essential that incrementing should always occur at a constant reference phase with respect to the carrier drive unless carrier drive is removed. If this is not done the size of the incremental flux change will be dependent on the vector sum of the switching and carrier signals. A typical scheme for realizing integrator operation is shown in Figure 8.

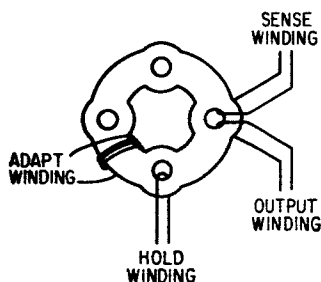


Figure 7 Multiple Aperture Device (MAD)

The physical requirement of providing a number of hand wound turns about the various apertures dictates to a large extent the cost of the device. Large driving currents, a moderate amount of timing during incrementing and relatively low output signal amplitude necessitate peripheral circuitry of considerable complexity. The resultant degradation in the basic reliability of the approach then becomes an imposing problem.

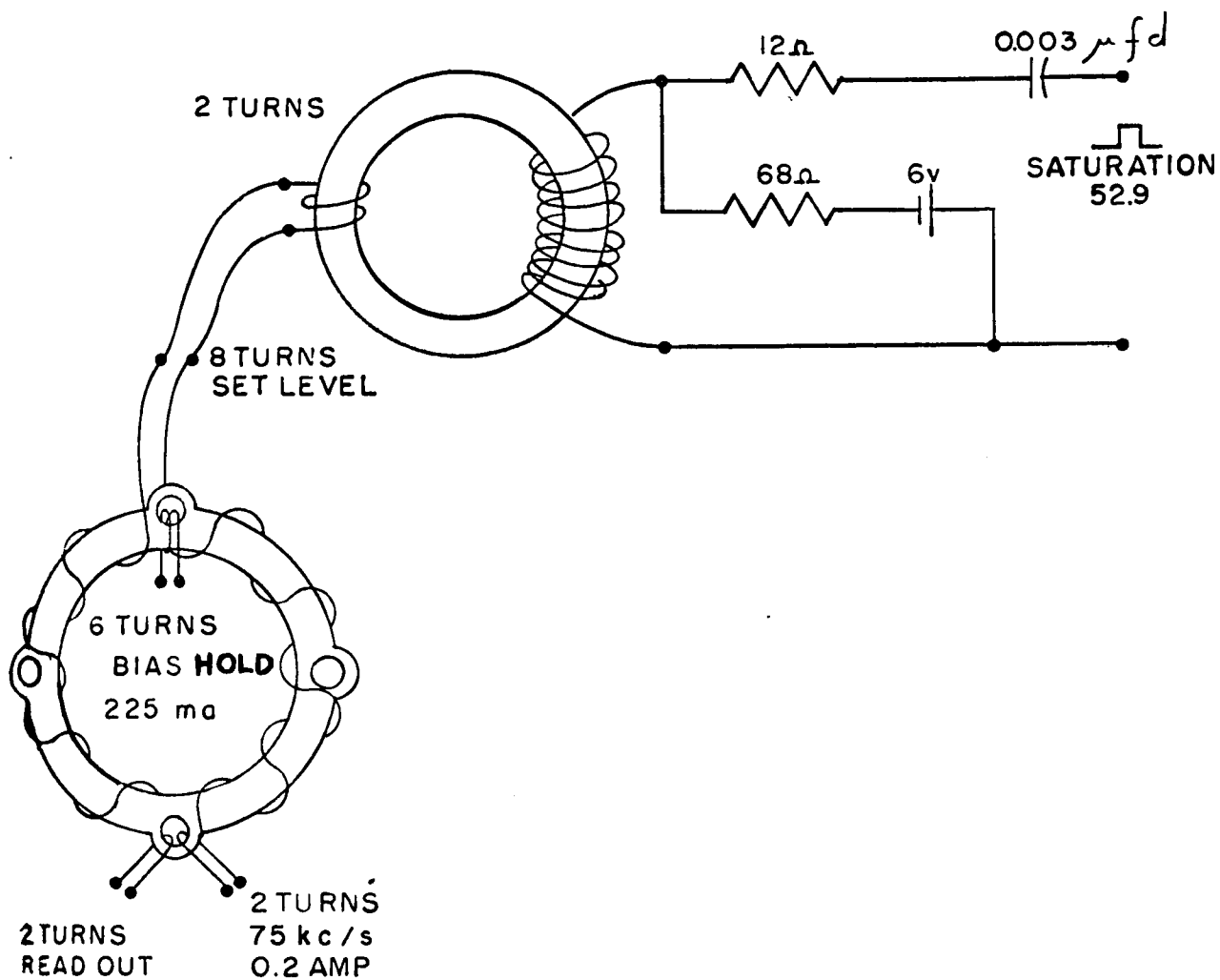


Figure 8 MAD Integrator

b. Orthogonal Core Integrator

The magnitude and direction of a stored flux may be sensed by applying a magnetic field orthogonally to the direction of stored flux.⁹ This causes the remanent flux vector to rotate generating a voltage proportional to its rate of change and hence its magnitude. The application of a read or sensing field at right angles to the stored or written flux minimizes the interaction of the sense drive on the stored flux magnetic path. At the termination of the read drive the flux vector returns back to its original preferred orientation by virtue of domain elasticity. A typical orthogonal core configuration is shown in Figure 9. The flux level stored in the core is altered by pulsing the output winding in a manner similar to the incrementing techniques previously discussed. Output signal consists of either positive or negative pulses depending upon the direction of the stored flux, with an amplitude proportional to the magnitude of the remanent flux. Practical problems similar to those associated with the multiaperture device previously discussed again make physical implementation cumbersome.

c. Second Harmonic Integrator¹⁰

Nondestructive readout of remanent flux may be obtained by reducing the sensing drive to a value insufficient to cause irreversible switching. Since magnetic cores are generally non-linear the output voltage will contain harmonics of the drive current. In particular, the even harmonic

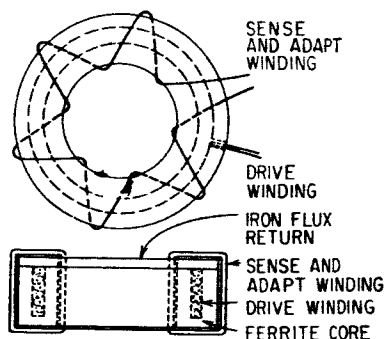


Figure 9 Orthogonal Core

voltage for certain core materials is found to be proportional to the net remanent flux level. The second-harmonic generator shown in Figure 10 consists of a pair of tape wound cores driven from an r-f sinusoidal power source. The output winding is arranged so that the fundamental component of drive voltage cancels out, leaving a second harmonic distortion voltage proportional to the remanent flux in the cores.

By passing a direct current through the output winding the remanent flux level may be altered. Due to an interaction between the d-c adapt current and the RF drive the rate of change of the remanent flux with respect to the adapt current is constant and reversible. Tape-wound cores have been found to provide the best performance and because of their higher permeability require fewer turns. Typical associated driving, sensing and timing circuitry tend to be rather elaborate however. The cancellation of the fundamental driving frequency is difficult to achieve in practice thus making the desired output signal appear against a background of noise. This low level signal must in turn be amplified in order to provide a signal compatible with the associated solid state circuitry which it must ultimately control. Clearly a separately switched driving source for each pair of cores is required in order to provide the individual binary signal inputs whose weights are to be altered. Since the sinusoidal drive currents tend to be in the order of 10 to 100 or more milliamperes the driving and peripheral circuitry is necessarily elaborate.

d. Magnetostrictive Integrator

The direction and magnitude of the net remanent flux in a magnetostrictive core may be sensed if the core is excited mechanically.¹¹ Figure 11 shows a simplified scheme for implementing a magnetostrictive storage system using an ultrasonic delay line to excite several magnetostrictive torroids. Driving source for the sonic delay line is a piezoelectric transducer. Input to each of the torroids is provided by means of narrow width

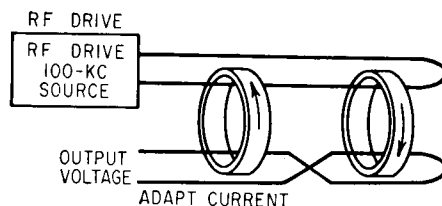


Figure 10 Second Harmonic Integrator

pulses through a separate write coil wound concentrically with the read coil. If the frequency and rms amplitude of the stress wave is maintained at constant value, the open circuit output of the read coil is approximately proportional to the flux stored in the individual torroids. Although this effect has been demonstrated experimentally by Nagy¹¹ and others the basic peculiarities of magnetic domain behavior especially under the influence of mechanical excitation is only crudely understood.

The experimental systems fabricated to date are rather large owing to the structural requirements of acoustical devices and the associated electronic circuitry necessary to provide proper timing, current driving and voltage amplification. At best considerable experimental work is necessary to show that magnetostrictive storage offers any real advantage over more conventional electro-magnetic approaches. Indeed, the sensing of remanent flux by acoustical means rather than by non-destructive, electrical drive appears to inject an unwarranted interface complexity.

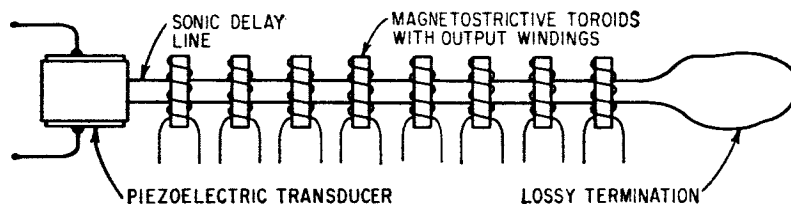


Figure 11 Magnetostrictive Integrator

3. Conclusion

As a result of the foregoing survey it became apparent that none of the suggested adaptive devices were sufficiently developed to justify the selection of a practical approach for immediate circuit implementation of an adaptive voter. An explicit evaluation was not attempted owing to the superficial treatment of the various devices by academic researchers.

The magnetic devices with their known sensitivity to temperature stress appear to offer the least hope for providing analog memory with long term stability. The requirement for providing carefully controlled incrementing with relatively large drive currents coupled with the small output signals and associated amplification appears to dictate an imposing amount of peripheral circuitry. The degradation in reliability as a result of this complexity represents a liability which makes practical application doubtful for redundant systems.

References

- 1) B. Widrow and M. E. Hoff, "Adaptive Switching Circuits," Technical Report No. 1553-1, Stanford Electronics Laboratories, June 1960.
- 2) B. Widrow, "Adaptive Sampled-Data Systems - A Statistical Theory of Adaption," 1959 WESCON Convention Record, part 4.
- 3) B. Widrow, "An Adaptive 'Adaline' Neuron Using Chemical Memistors," Technical Report No. 1553-2, Stanford Electronics Laboratories, October 1960.
- 4) "An Introduction to Solions," Texas Research and Electronic Corp., Dallas, June 1961.
- 5) "D-C Amplifier Uses Fluid-State Tetrode," Electronic Products Magazine, October 1962.
- 6) "Capacitive Readout Integrator," Technical Brochure, Curtis Instruments, Inc., Mount Kisco, New York.
- 7) J. A. Rajchman and A. W. Lo, "The Tranfluxor," Proceedings of the I.R.E., March 1956.
- 8) A. E. Brain, "The Simulation of Neural Elements by Electrical Networks based on Multi-Aperture Magnetic Cores," Proceedings of the I.R.E., January 1961.
- 9) J. K. Hawkins and C. J. Munsey, "A Magnetic Integrator for the Perceptron Program," Annual Summary Report, Publication No. U-603, Aeronautics, Newport Beach, Col., July 30, 1960.
- 10) H. S. Crafts, "A Magnetic Variable Gain Component for Adaptive Networks," SEL-62-1147, Technical Report 1851-2, Stanford Electronics Laboratories, December 1962.
- 11) G. Nagy "Analogue Memory Mechanisms for Neural Nets," Cognitive Systems Research Program, Contract No. NONR 401(40), Report No. 3, Cornell University, Ithaca, New York, August 31, 1962.

Appendix 4

TRANSOR ANALYSIS

by

R. S. Bray

P. A. Jensen

C. G. Masters

September 1963

TABLE OF CONTENTS

I.	INTRODUCTION	4-1
II.	RESTORING CIRCUIT MODELS	4-3
	A. The Transor Decision Function	4-3
	B. The Threshold Decision Function	4-4
III.	FAILURE MODES	4-6
	A. Transor Restoring Circuit Vulnerability	4-6
	B. Threshold Restoring Circuit Vulnerability	4-10
IV.	RELIABILITY ANALYSIS	4-11
	A. Transor Reliability Defined	4-11
	B. Output Modes Defined	4-11
	C. Upper Bound on Transor Reliability	4-12
	D. Transor Reliability for Strictly Asymmetric Failure Modes	4-13
	E. Transor Reliability for Mutually Exclusive Output Failure Modes	4-13
	F. Transor Reliability for Symmetrical Environment	4-15
V.	CONCLUSION	4-17
	BIBLIOGRAPHY	4-25

I. INTRODUCTION

In recent years many novel schemes have been proposed to improve digital system reliability through the use of "redundant" equipment. Several of these, patterned after a concept of Von Neumann,¹ require a "restoring organ," "restorer" or "voter" to be placed after each set of redundant signal processors which perform a particular subsystem function. A restoring organ receives an input from each member of the associated set of processors. From these nominally identical input signals, the restoring organ produces an estimate of the correct subsystem output based on one or more specified decision criteria. It should be noted that the restorer does not perform any data processor function but acts as an error correcting transmission channel connecting two signal processors.

It has been shown in the literature² that the theoretically most efficient restoring organ is one that is capable of adapting itself to changes in the reliability of inputs. Specifically, for threshold type organs it has been shown that the optimum use of n unreliable versions of the same signal could be achieved by dynamically weighting each input in accordance with its relative reliability. Inputs which have a past history of being more reliable are given the heavier vote weights, and the unreliable inputs the lighter vote weights. The ideal restoring organ would sense the unreliable inputs and decide on the optimal vote weights. By efficiently tailoring the restoring organ to its ever-changing environment, significant improvement could be achieved over the presently popular majority restoring circuits.

In studying adaptive restoring organs, Company A has shown³ that circuit implementation of adaptive restoring organs for the specific requirements of redundant space-borne systems is not yet practical. The complex circuitry required under the present "state of the art" to perform the adaptive function results in machines too cumbersome and unreliable to compete with less sophisticated redundant systems. This does not mean though that the present restoring organs used in redundancy techniques are adequate and cannot be improved upon.

The purpose of this study is to investigate a new restoring organ proposed by Company A, called the Transor⁴. A characteristic of many failed subsystems is their tendency to have steady-state outputs as their dominant failure mode. In Transor, steady-state outputs are automatically deweighted by detecting only changes in states rather than the absolute states themselves. In an environment where the probability of steady state

^{1, 2, 3, 4} See Bibliography

failure is relatively high, a restoring organ which ignores its steady-state inputs can derive a correct output with less than a majority of working inputs.

The salient characteristics of the Transor restoring organ are best shown by contrasting them to the corresponding characteristics of a majority restoring organ. The majority organ was chosen as a reference base because of its similarity in function to the Transor and because it is presently the most widely used restoring organ.

II. RESTORING CIRCUIT MODLES

A. THE TRANSOR DECISION FUNCTION

To be consistent with the terminology adopted by one group of investigators, the term "restoring circuit" will be used to denote one functional unit of a restoring organ or restorer. A very general block diagram of a Transor restoring circuit having binary inputs (x_1, x_2, \dots, x_R) and an output z is shown in figure T-1.

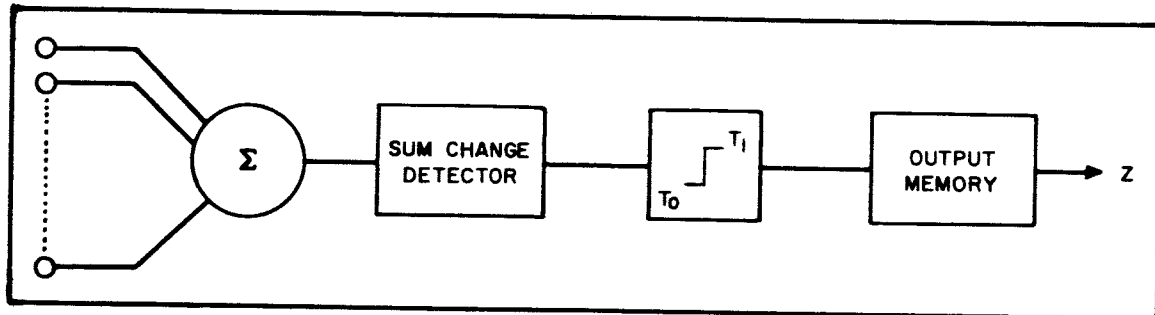


Figure T-1. Transor Restoring Circuit

Some of the salient characteristics of a Transor Restoring circuit are noted below:

- 1) It has memory
- 2) It operates only on the number of changes in the states of individual inputs between two adjacent bit times, $(t - 1)$ and (t) .
- 3) It is a binary voting element with a binary output.
- 4) It has two thresholds, not necessarily of the same magnitude, which combine with the states of the input at $(t - 1)$ and (t) to determine the element output.

The functional relationship, describing the Transor Decision function can be stated as follows

$$z^{(t)} = f \left[z^{(t-1)}; (x_1, x_2, \dots, x_R)^{(t)}; (x_1, x_2, \dots, x_R)^{(t-1)}; T_0; T_1 \right] \quad (1)$$

The number of binary Ones appearing on its inputs during each bit time are summed and compared with the number present during the previous time period. If the change is positive and greater than a given threshold T_1 then the output z is forced to a binary One. If the change is negative and greater in magnitude than a second threshold, T_0 , the output is forced to a binary Zero. If neither threshold is exceeded, the output does not change from its previous state. This operation may be summarized by the following decision rule statements.

$$\sum_0^R x_i^{(t)} - \sum_0^R x_i^{(t-1)} \geq T_1 \rightarrow Z^{(t)} = 1 \quad (2)$$

$$\sum_0^R x_i^{(t)} - \sum_0^R x_i^{(t-1)} \leq -T_0 \rightarrow Z^{(t)} = 0 \quad (3)$$

$$-T_0 < \sum_0^R x_i^{(t)} - \sum_0^R x_i^{(t-1)} < T_1 \rightarrow Z^{(t)} = Z^{(t-1)} \quad (4)$$

B. THE THRESHOLD DECISION FUNCTION

The threshold model* consists of a black box having a certain number of binary inputs (x_1, x_2, \dots, x_R) and an output z . At any bit time (t) the state of the output line z is a function of the state of the input lines and the threshold T . A general relationship similar to equation (1), but describing the threshold decision function may be delineated by the following expression.

$$Z^{(t)} = f \left[\left(x_1, x_2, \dots, x_R \right)^{(t)} ; T \right] \quad (5)$$

If the output, z , can assume either a Zero or One state, the threshold restoring circuit makes a decision to force its output to the One state under the following decision rule:

* The majority gate is a threshold model with $T = \frac{R+1}{2}$, where R is the number of inputs.

If

$$\sum_{0}^R x_i^{(t)} \geq T \rightarrow Z^{(t)} = 1 \quad (6)$$

and to the Zero state when

$$\sum_{0}^R x_i^{(t)} < T \rightarrow Z^{(t)} = 0 \quad (7)$$

III. FAILURE MODES

A. TRANSOR RESTORING CIRCUIT VULNERABILITY

Before the reliability of any Transor network can be expressed in a meaningful mathematical form, the failure modes of the individual subsystems appearing at the Transor's inputs must be explicitly stated.

A characteristic of Transor is its ability to differentiate between transistional and steady-state failures. This property creates failure modes different from those of threshold decision. Specifically, a signal processor is assumed either to be working correctly or failed into one of the following modes:

1. The transitional mode, in which extra Ones and/or extra Zeros appear at the output, and
2. The steady-state mode, in which the output permanently remains in a single state.

A transition (figure T-2) is defined as the rise or fall of a pulse during its switching time. The restoring circuit executes a decision by vector summing the change

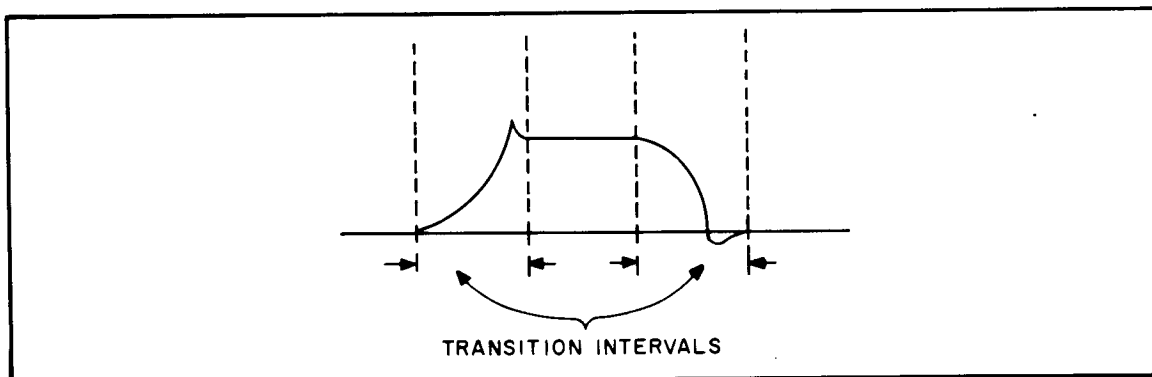


Figure T-2. Transition Intervals

in input pulses on the R redundant lines during the vote interval and a decision is made according to the decision rules (2) through (4). The term "extra One" implies a one has appeared on a signal processor's output when it should have been a Zero. By going to the wrong state a signal processor creates a wrong transition which is voted by the Transor. Wrong transitions can occur through diode failures in the gating section of diode-transistor

type signal processors. These failures sporadically generate "extra" Ones or "extra" Zeros as a function of the information at the gate's inputs. To illustrate, consider a three input Transor voting on the output of a network of redundant AND gates. The state of the binary inputs may be represented by the state vector $S_i^{(t)}$ below.

$$S_i^{(t)} = \begin{bmatrix} x_1^{(t)} \\ x_2^{(t)} \\ x_3^{(t)} \end{bmatrix}$$

In figure T-3 a diode is assumed to have opened in branch (1) of two of the gates causing those branches to appear as Ones. An erroneous One will appear at the gate's output whenever a correct Zero appears on those inputs and correct Ones appear on the remainder of the inputs. However, if all the input diodes open or an output element fails, the gating function will be destroyed, and the output will assume a steady-state. A method for determining the probability that a signal processor will fail into either of these two modes is discussed in Appendix I.

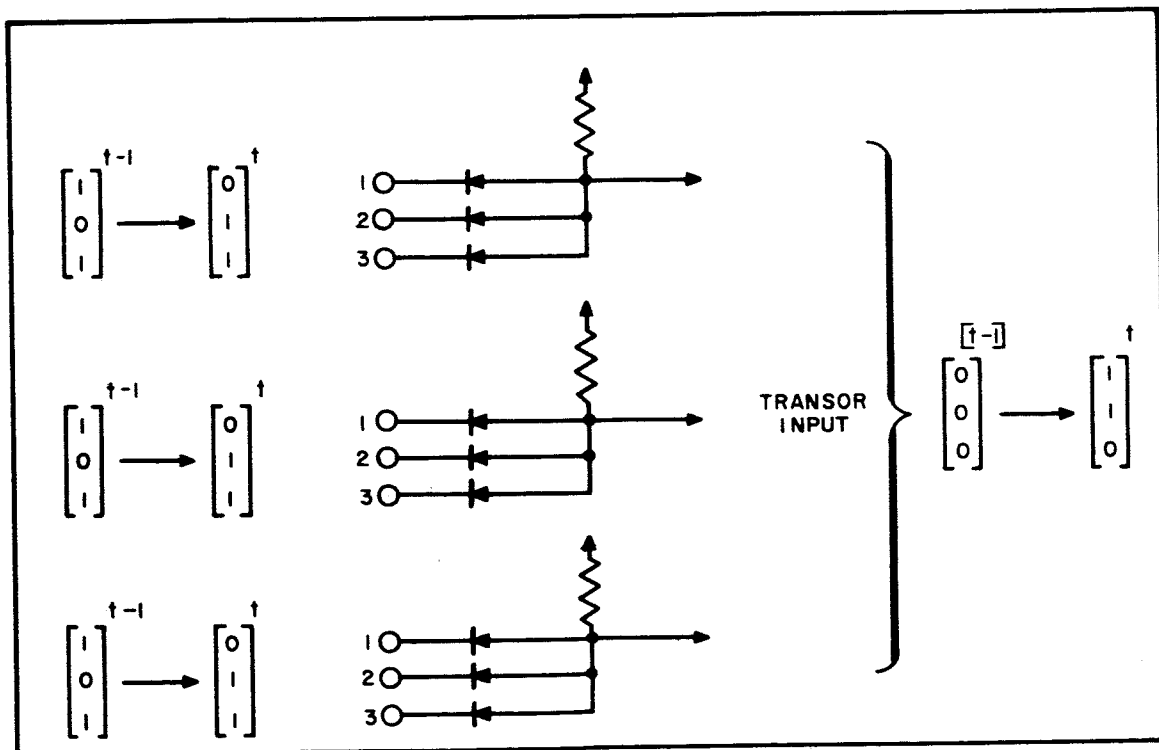


Figure T-3. Generation of Wrong Transitions in Redundant AND Gates

Because transitions are vector quantities their occurrence in the wrong direction may threaten Transor performance in three ways:

1. Wrong transitions cancelling correct transitions.
2. Wrong transitions occurring while the correct inputs remain in the same state (a series of Ones or Zeros). During this time the correct inputs have lost their voting power.
3. Wrong transistions temporarily simulating steady-state failures.

Wrong transitions produced by "extra Ones" and/or "extra Zeros" over a sequence of bit times can result in "error correlation" and create a variety of failure modes, subject to the nominally correct input states to the Transor for the considered sequence.

Figure T-4 shows this more clearly when state vectors are used to represent the inputs to a five input Transor. Inputs x_1 and x_2 are assumed to have failed and capable of randomly producing wrong transitions in either direction, i.e., extra Ones or Zeros. No inputs are assumed failed to a steady-state. For definiteness all inputs at time (t) may be assumed correct. In the following bit times (proceeding to the right) several failure patterns are possible for each nominally correct input state. At (t+1) the states (2), (3), (4), and (5) are considered among the possible states (four other possible states including (1) have been omitted as repetitious). Observe that sequence (1) \rightarrow (2) is the most damaging because only the wrong transitions have any voting power. For a threshold set as low as two this would result in a wrong decision. The sequence (1) \rightarrow (5) represents a possibility in which both erroneous inputs have temporarily "stuck" in one state simulating a temporary steady-state. The sequences (1) \rightarrow (3) and (1) \rightarrow (4) are the most likely possibilities in which one of the failed inputs is temporarily correct. In the next bit time (t + 2) transitions to the possible states (3), (4), (5) and (6) and (7) are considered (again repetitions are omitted). Shown here are the cancellation effects caused by the introduction of errors on the previous bit time, demonstrating the "error correlation" inherent in Transor. The sequence (2) \rightarrow (5) is the most damaging because any threshold greater than one would have resulted in a wrong decision. Observe the tradeoff conflict created by the necessity for setting the threshold at a value greater than two in the sequence (1) \rightarrow (2) and the same threshold at a value less than two in the sequence (2) \rightarrow (5) in the following bit time. Clearly there must exist an optimum threshold. Inclusion in figure (4) of transitions from states (4) and (5) would have produced no new failure modes since they are but the duals of (2) and (3).

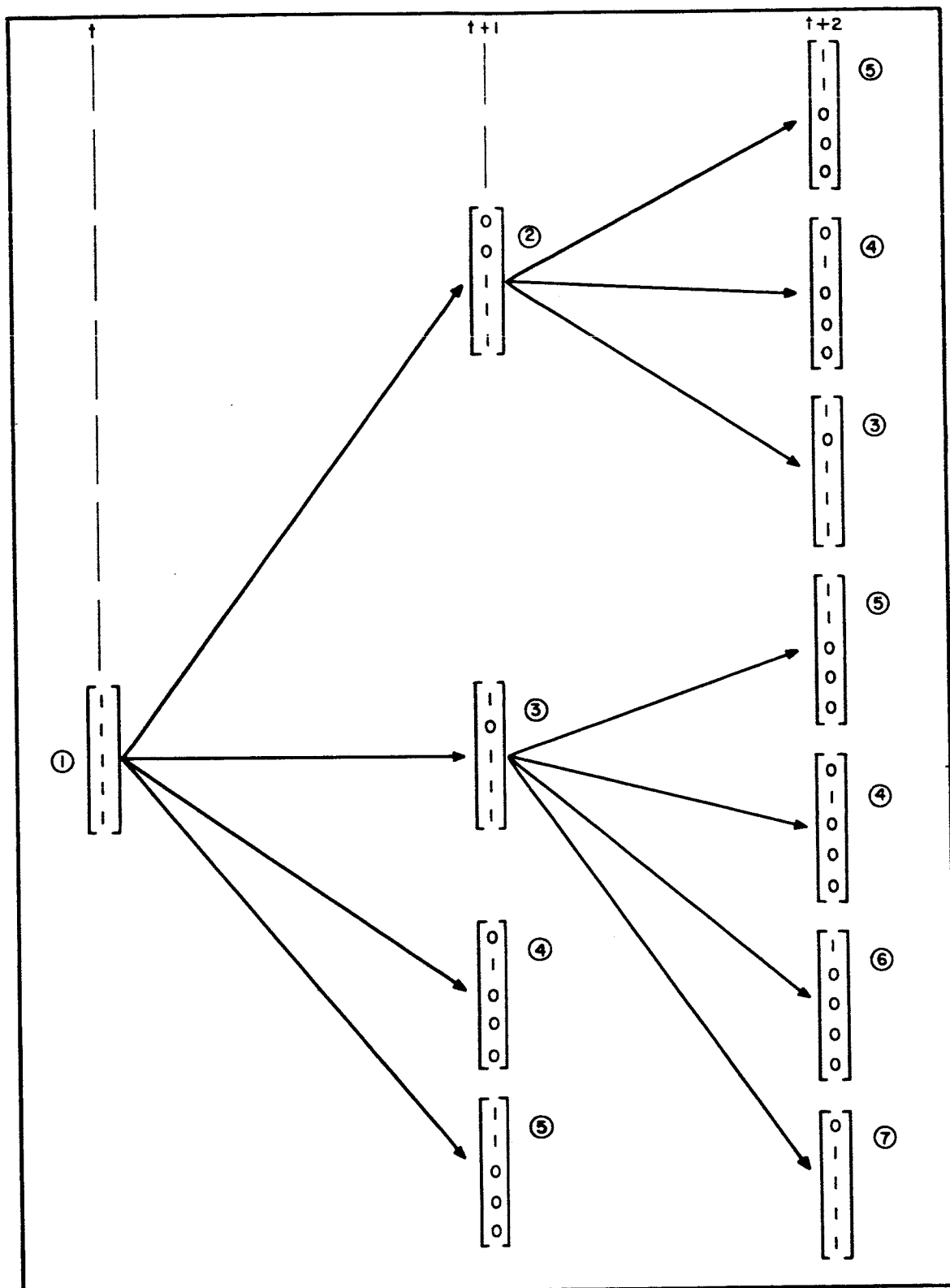


Figure T-4. Possible Sequences of Input States for a Five Input Transor Over Two Bit Times

B. THRESHOLD RESTORING CIRCUIT VULNERABILITY

A threshold restoring circuit makes a decision at time (t) by summing the number of binary ones appearing momentarily at its inputs. The decision is independent of the input state at time (t - 1). By virtue of decision rule (6) if the number of errors appearing on the restorer's inputs is greater than the threshold T the restorer makes the wrong output decision. As opposed to Transor, the threshold device cannot differentiate between pure wrong transitions and steady-state failures so that both failure modes may be lumped together. To illustrate, consider a three-input threshold restoring circuit whose threshold is set at two (T = 2). For definiteness assume that x_1 and x_2 at time (t) are in error and in the same state and x_3 is correct as indicated below.

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \bar{x} \\ x \end{bmatrix} \rightarrow Z(t) = \bar{x}$$

Under this condition a wrong decision will be made. This may be considered a "worst case" failure mode because the alternate situation is possible where x_1 and x_2 have failed into opposite steady states.

$$\begin{bmatrix} 1 \\ 0 \\ x_3 \end{bmatrix} \rightarrow Z = x_3$$

In this case the errors nullify each other and the restoring circuit's output will follow the single correct input (x_3). In most reliability analyses the "worst case" is assumed, and any two failures in a set of restoring circuit inputs are assumed to cause system failure.

IV - RELIABILITY ANALYSIS

A. RELIABILITY DEFINED

In keeping with the usual concept of reliability, the reliability of a Transor restoring circuit will be defined as the probability that it never makes a wrong decision during its mission time. For analysis purposes the transor itself is assumed perfectly reliable, i. e., a wrong decision is never made through component failure within the Transor itself. In part III it was shown that errors appearing on the Transor inputs in a particular bit time could be correlated with errors that appeared on adjacent bit times to produce unique failure modes. Two of these were:

- (1) Cancellation effects
- (2) Simulated steady-state

In the following discussion it will be shown how these failure modes may be "built in" to reliability models by using multinomial expansions. Analytical models formulated in this manner may be easily compared with models for threshold reliability.

B. OUTPUT MODES DEFINED

Any output of a binary signal processor can be classified into one of six mutually exclusive classes over the element's mission time. These are:

- 1) Correct
- 2) Continuous Zero state
- 3) Continuous One state
- 4) Extra Ones but no extra Zeros
- 5) Extra Zeros but no extra Ones
- 6) Both extra Ones and Zeros.

Moreover the output of a system, composed of binary signal processors may be defined by the six mutually exclusive classes above. Each of these classes will be assigned the following probability measures in conformance with the Transor decision rules.

- 1) p ; the probability that the output is correct
- 2&3) q_s ; the probability that the output is either a continuous Zero or a continuous One.
- 4) q_1 ; the probability that the output generates extra Ones, but not extra Zeros.

- 5) q_0 ; the probability that the output generates extra Zeros, but not extra Ones
- 6) q_{10} ; the probability that the output generates both extra Ones and Zeros randomly.

Note that the measure q_s is the result of the union of classes (2) and (3). The transitional probabilities q_1 , q_0 and q_{10} are defined to represent only the probabilities that a particular set of components, whose failure will cause wrong transitions to be generated randomly, will fail.

C. UPPER BOUND ON TRANSOR RELIABILITY

An upper bound on reliability is easily obtained by excluding all but steady-state failures from the environment. If β is a random variable denoting the number of correct transitions (or working inputs) and γ the number of inputs failed to a steady-state; a probability density function may be defined over the sample space as

$$\vartheta = \binom{R}{\beta, \gamma} P^\beta q_s^\gamma$$

Since Transor ignores steady-state failures the only criterion for a correct decision is that

$$\beta \geq T_0$$

$$\beta \geq T_1$$

The corresponding limits on γ are

$$\gamma \leq R - T \tag{8}$$

where $T_0 = T_1 = T$. The reliability is

$$R_{U. B.} = \sum_{\beta = T}^R \binom{R}{\beta} P^\beta (1 - P)^{R - \beta} \tag{9}$$

In an environment capable of producing only steady-state failures, the maximum reliability and error correction capability is obtained by setting $T = 1$. This is the optimum threshold. From equation (8) we see that Transor can correct at best $R - 1$ failures in an order R redundant system.

D. TRANSOR RELIABILITY FOR STRICTLY ASYMMETRIC FAILURE MODES

Excluding from the mutually exclusive ways an environment can fail class (6) and either class (4) or (5) limits transitional failure modes to states (2), (3), (4) and (5) in fig. (4). Of these the sequence (1) → (2) is the "worst case". For definiteness let it be assumed that Transor inputs may produce only extra Zeros and steady-state failures. Let α_0 be a random variable denoting the number of wrong transitions to the Zero state.

The density function on this sample space is

$$\theta = \binom{R}{\beta, \gamma, \alpha_0} P^\beta q_s^\gamma q_0^{\alpha_0}$$

A wrong decision will be made unless

$$\alpha_0 \leq T_0 - 1$$

Since it is necessary that

$$\beta \geq T_0$$

the limits on γ must be

$$\gamma \leq R - T_0 - \alpha_0$$

The reliability is

$$R = \sum_{\alpha_0=0}^{T_0-1} \sum_{\gamma=0}^{R-T_0-\alpha_0} \binom{R}{R-\alpha_0-\gamma, \gamma, \alpha_0} P^{R-\alpha_0-\gamma} q_s^\gamma q_0^{\alpha_0} \quad (10)$$

E. TRANSOR RELIABILITY FOR MUTUALLY EXCLUSIVE OUTPUT FAILURE MODES

The scope of the environment considered in part D can be broadened to include both the mutually exclusive classes (4) and (5). Each input may be failed to either steady-state, extra Ones or extra Zeros (but not both). The failure modes (figure T-5) may be represented in a manner similar to figure T-4; inputs x_1 and x_2 assumed failed in one of the four mutually exclusive ways listed above.

The sample space may be described by the density function

$$\theta = \binom{R}{\beta, \alpha_0, \alpha_1, \gamma} P^\beta q_0^{\alpha_0} q_1^{\alpha_1} q_s^\gamma$$

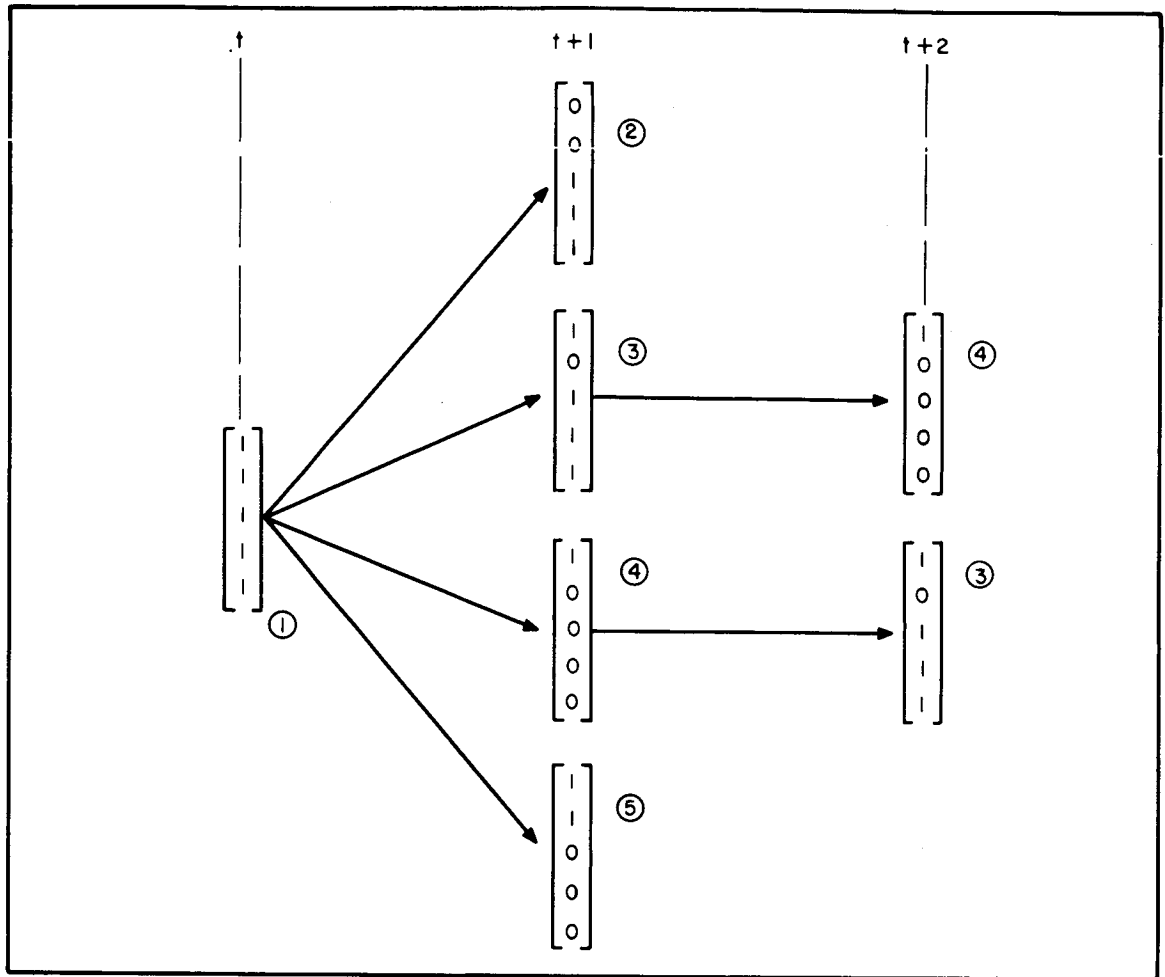


Figure T-5. Possible Sequences for a Five-Input Transor with Mutually Exclusive Output Failure Modes

The sequence (1) \rightarrow (2) in figure (5) implies that a Transor will make a wrong decision unless

$$\alpha_0 \leq T_0 - 1 \quad (11)$$

and its dual

$$\alpha_1 \leq T_1 - 1. \quad (12)$$

From the sequences (1) \rightarrow (3) and (1) \rightarrow (4) respectively

$$\beta + \alpha_1 \geq T_1 \quad (13)$$

$$\beta + \alpha_0 \geq T_0 \quad (14)$$

for a correct decision. However examination of the sequences (3) — (4) and (4) — (3) show that inequalities (13) and (14) do not represent "worst cases". "Error correlation" between the bit times (t + 1) and (t + 2) have produced a temporary steady-state. A correct decision will be made only if

$$\beta \geq T_0 \quad (15)$$

$$\beta \geq T_1 \quad (16)$$

From (15) and (16)

$$\gamma \leq (R - T_0) - \alpha_1 - \alpha_0 \quad (17)$$

$$\gamma \leq (R - T_1) - \alpha_1 - \alpha_0 \quad (18)$$

Of these last two inequalities the number of allowable steady-state failures, γ , will be governed by the highest threshold, T_0 or T_1 .

The reliability will take the form

$$R = \sum_{\alpha_0=0}^{T_0-1} \sum_{\alpha_1=0}^{T_1-1} \sum_{\gamma=0}^{R-T_0-\alpha_0-\alpha_1} \left(\begin{matrix} R \\ R-\alpha_0-\alpha_1-\alpha, \alpha_0, \alpha_1, \gamma \end{matrix} \right)^p q_0^{\alpha_0} q_1^{\alpha_1} q_s^\gamma \quad (19)$$

where T_0 is assumed $> T_1$.

F. TRANSOR RELIABILITY FOR A SYMMETRICAL ENVIRONMENT

A symmetrical environment utilizing Transor decision will be defined as the mutually exclusive classes (1), (2), (3) and (6). Wrong transitions may occur in both directions and at random. Therefore $\alpha_0 = \alpha_1 = \alpha$ and $T_0 = T_1 = T$. The density function on this sample may be written as

$$\phi = \left(\begin{matrix} R \\ \beta, \alpha, \gamma \end{matrix} \right) p^\beta q_{10}^\alpha q_s^\gamma$$

From figure T-4 it can be seen that a wrong decision will be made unless

$$\alpha \leq T - 1 \quad (20)$$

and $\beta - \alpha \geq T \quad (21)$

From (21)

$$\gamma \leq R - T - 2\alpha \quad (22)$$

Therefore the reliability for the symmetrical environment is

$$R = \sum_{\alpha=0}^{T-1} \sum_{\gamma=0}^{R-T-2\alpha} \binom{R}{R-\alpha-\gamma, \alpha, \gamma} p^{R-\alpha-\gamma} q_{10}^{\alpha} q_s^{\gamma} \quad (23)$$

V. CONCLUSION

The dynamic characteristics of the Transor decision function make this type restoring circuit unique to the present art. The mission of this part of the Failure Free Systems Study has been to evaluate the potential usefulness of the Transor as a restoring circuit.

Primarily because it is most commonly used in present redundant equipment, the threshold type restoring circuit has been chosen as the reference point for the evaluation primarily. It has been hypothesized that, if it can be shown that the Transor failure masking capability compares favorably to that of the threshold restoring circuit, further development, including the construction of a breadboard model, should be justified.

The results of section IV have shown that there are certain environments in which Transor can be used to advantage in improving system reliability. For example, the maximum error restoring capability of Transor is shown to be $R-1$ failures of R redundant lines in an environment free from transitional failures. This is a significant improvement over the majority threshold restoring capability under the same conditions. There is need for caution, however, for in environments where symmetrical transitional errors are possible error correlation may make Transor performance inferior to threshold. From the reliability models, a tradeoff may be determined in terms of the output error probabilities of the environment.

The work done up to this point represents only a first step in Transor decision study. Work yet to be done includes: (1) a general Transor reliability model incorporating all the possible failure modes and (2) a decision rule for determining an optimum threshold.

In addition to continuing the analytical effort described in this report, a computer simulation program is being written to aid in the task (1) effort. This will be a relatively simple but versatile program designed to accommodate any set of restricting assumptions including those made in the four models derived in this report. The results of this report have shown a solution to task (2) would be desirable because of the tradeoffs between different failure modes. If the error probabilities of the signal processor outputs are known in the design stage maximum reliability can be bought for zero additional cost by a judicious choice of the thresholds.

VI. APPENDIX

Determination of the Reliability Parameters p , q_s , q_o , q_1 , q_{10} in a Signal Processor.

In section IV it was shown that reliability models could be formulated in terms of the output error probabilities of a set of redundant signal processors. This section describes a method for determining these probabilities.

Consider a set X^* which has for its members the n components of a signal processor. Each member (component) has two possible states:

x_i ; the i^{th} member is working.

\bar{x}_i ; the i^{th} member has failed.

Let each component have a reliability

$$P(x_i) = e^{-\lambda_i t}$$

and a probability of failure

$$P(\bar{x}_i) = 1 - e^{-\lambda_i t}$$

The probability measure on the sample space of X may be partitioned into the canonical form

$$\begin{aligned} 1 = & P(x_1 \cap x_2 \cap \dots \cap x_n) + P(\bar{x}_1 \cap x_2 \cap \dots \cap x_n) \\ & + P(x_1 \cap \bar{x}_2 \cap x_3 \cap \dots \cap x_n) + \dots + \\ & + P(\bar{x}_1 \cap \bar{x}_2 \cap \dots \cap \bar{x}_n) \end{aligned} \quad (24)$$

Briefly, the method requires determining the correspondence between groups of the terms in (24) and the individual terms in

$$1 = p + q_s + q_o + q_1 + q_{10} \quad (25)$$

Obviously the parameter p , that the signal processor output is correct is

$$p = P(x_1 \cap x_2 \cap \dots \cap x_n)$$

The remaining $2^n - 1$ terms in (24) are mapped into the four remaining parameters in (25) by partitioning the set X into subsets whose members are defined by those components whose

* Summary of all the notation to be used is included on the last page of this appendix.

failure will result in one of the four mutually exclusive events described in part IV. Specifically let

X_{ss} be the set whose failure results in either a steady-state Zero or One.

X_1 be the set whose failure results in extra Ones.

X_0 be the set whose failure results in extra Zeros.

X_{10} be the set whose failure results in extra Ones and Zeros.

Since each component may fail by shorting or opening, these two modes will determine membership in one or more of the above sets. If the probability of a component shorting given that its failed, $P(x_i^s | \bar{x}_i)$, is ρ_i then the joint probability of x_i failing and shorting is

$$P(\bar{x}_i \cap x_i^s) = P(x_i^s) = P_i(1 - e^{-\lambda_i t})$$

Let the probability of an x_i opening given that its failed the $P(x_i^o | \bar{x}_i)$

Then

$$P(x_i^s | \bar{x}_i) + P(x_i^o | \bar{x}_i) = 1$$

and

$$P(x_i^o | \bar{x}_i) = 1 - \rho_i$$

Also since for each x_i the events working, shorted or opened are mutually exclusive the probability of a component not shorting is

$$P(\bar{x}_i^s) = P(x_i \cup x_i^o) = 1 - \rho_i(1 - e^{-\lambda_i t})$$

To illustrate the technique a NAND gate will be analyzed using the test results contained in an earlier report.⁵

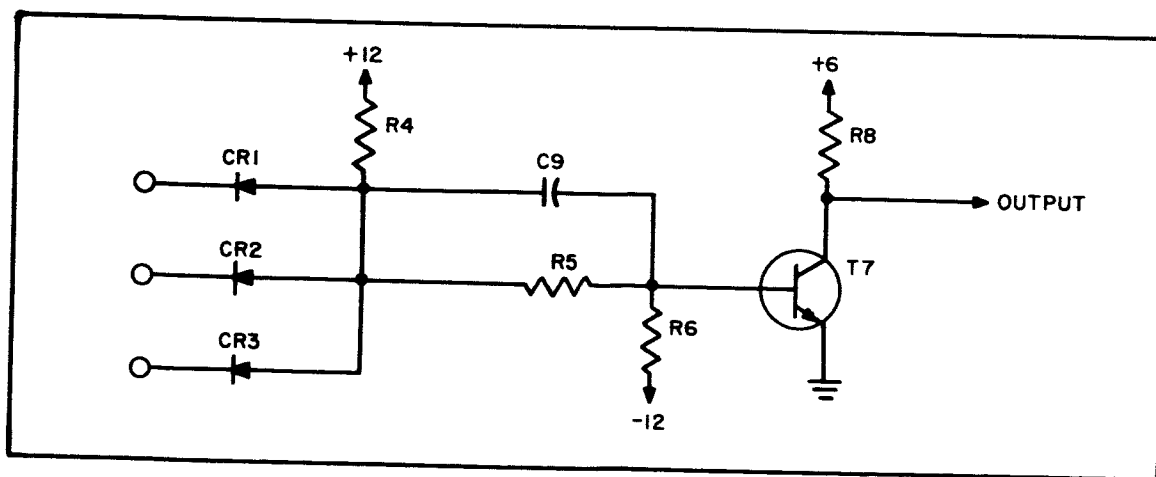


Figure AT-1. NAND GATE

The pertinent results are included below.

1. AND gate input diodes; CR1, CR2, CR3

A. OPEN - Any open circuit input is equivalent to a logical "one" on that input; it cannot inhibit the AND gate.

B. SHORT - A shorted diode will not affect the ability to perform the AND function if that input has low impedance to ground in the "zero" state and high impedance to a positive voltage in the "one" state. The line with a shorted diode is no longer isolated from other inputs; that line is shorted to the AND gate output and may, therefore, be an incorrect "zero".

2. AND gate resistor; R4

A. OPEN - The AND gate has no voltage available to drive current into the transistor base, so the NAND gate output remains a "one".

B. SHORT- This will cause a low impedance path from the +12 volt power supply through the input diodes to all of the inputs to the gate. If any of these inputs are from NAND gate transistors which are conducting, that input will also be a low impedance to ground. A low impedance path then exists from the power supply to ground, and a high current will flow through the diode and transistor according to the magnitude of the impedance of the power supply and components involved. In the tests observed, this current was not sufficient to damage the transistor or diode and did not blow the fuse on the power supply. However, if any inputs are from flip-flops, the clamp diode will turn on when the voltage exceeds the clamp voltage. A low impedance path then exists from the +12 volt power supply through the shorted AND gate resistor, the input diode, and may seriously overload the clamp voltage supply, depending how the clamp voltage is derived. In the tests observed, this current was sufficient to cause both the input diode and clamp diode to short and the clamp voltage to rise toward +12 volts.

3. Input resistor - capacitor; R5, C9

A. Resistor SHORT- The transistor base voltage will be the AND gate output. This will normally cause the transistor to conduct, so that the output will be "zero" for any logic input.

B. Resistor OPEN- This will cause the transistor to be off, so that the output will be a "one" for all logic inputs.

C. OPEN C9 - This does not adversely affect operation, unless the switching time is critical, in which case NAND gate turn-on time was increased from 65 nanoseconds with C9 to 80 nanoseconds without C9; turn-off time was increased from 25 to 45 nanoseconds in one approximate measurement with a constant load on the output of the circuit. The turn-on time was measured as the time from the input going positive above +1.6 v. until the output goes to +1.6 v. from the "one" state. The turn-off time was measured as the time from the input going negative below +2.4 v. until the output goes to +2.4 v. from the "zero" state.

4. Base bias resistor, R6

- A. OPEN - This will normally cause the transistor to conduct, so that the output will be "zero" for any logic input, except that when the AND gate voltage is going negative from the "one" state, this voltage change is coupled across C9 and will turn the transistor off until the transient effect has ended.
- B. SHORT- The short of the base resistor may cause damage to the output transistor, since -12 volts on the base exceeds the maximum rating of 5 volts for V_{EBO} . The output voltage will depend on the failure mode, if any, of the transistor. In three multiple failure tests that included short of the base bias resistor in a NAND gate, two transistors shorted base to collector, which resulted in a -12 volt output; one transistor shorted collector to emitter, which resulted in a "zero" output. The -12 volt output did not cause any significant difference than a normal "zero" output to the following circuitry.

5. Collector (output) resistor, R8.

- A. OPEN- The removal of the output resistor does not affect the logical operation of the circuit, since any loads are also to positive voltage sources. The output rise time will be somewhat slower but the output will turn off faster because the output voltage in the "one" state is lower and the load current is less.
- B. SHORT- The output voltage will be +6 volts; the current in the transistor will be high if the transistor is conducting. This current was not sufficient to cause permanent damage to the transistor in the observed tests.

6. Transistor, T7

The transistor may fail into any of several possible modes, but the circuit output will usually be a "one" unless a low impedance path exists from the output to ground, such as when the collector is shorted to emitter, or if the transistor is otherwise forced to remain conducting from collector to emitter.

From the test results the component failures may be categorized (below) into their effects on the NAND gate's output.

I Components Causing Failure into Steady State "1"

- 1) R4 Open
- 2) R5 Open
- 3) T7 (most modes result in a "1")

II Components Causing Failures into Steady State "0"

- 1) R5 short
- 2) R6 short
- 3) R6 open
- 4) CR1 and CR2 and CR3 open (together)

III Component Failures that will Produce Transitional Extra "Ones"

- 1) CR1 or CR2 or CR3 open
- 2) CR1 and CR2 open
- 3) CR1 and CR3 open
- 4) CR2 and CR3 open

From the three categories above may be formed the mutually exclusive sets

Set X_s	Probability of $X_s (i) = P [X_s (i)]$
$X_s (1): x_4^0$	$(1 - \rho_4) (1 - e^{-\lambda_4 t})$
$X_s (2): \bar{x}_5$	$1 - e^{-\lambda_5 t}$
$X_s (3): \bar{x}_6$	$1 - e^{-\lambda_6 t}$
$X_s (4): \bar{x}_7$	$1 - e^{-\lambda_7 t}$
$X_s (5): x_1^0 \cap x_2^0 \cap x_3^0$	$\left[(1 - \rho) (1 - e^{-\lambda t}) \right]^3$

The probability of a steady-state failure is

$$q_s = \sum_{i=1}^5 P [X_s (i)] - \sum_{i \neq j}^5 P [X_s (i, j)] + \sum_{i \neq j \neq k}^5 P [X_s (i, j, k)] \\ - \sum_{i \neq j \neq k \neq l}^5 P [X_s (i, j, k, l)] + \prod_{i=1}^5 P [X_s (i)]$$

Set X_0

$$X_0(1): (x_1^0 \oplus x_2^0 \oplus x_3^0) \cap \bar{x}_4^0 \cap x_5 \cap x_6 \cap x_7$$

$$X_0(2): (x_1^0 \cap x_2^0 \oplus x_2^0 \cap x_3^0 \oplus x_1^0 \cap x_3^0) \\ \cap \bar{x}_4^0 \cap x_5 \cap x_6 \cap x_7$$

Probability of $X_0(i) = P[X_0(i)]$

$$3(1-e)(1-e^{-\lambda t}) \cdot$$

$$e^{-2\lambda t} \left[1 - (1-\rho_4)(1-e^{-\lambda_4 t}) \right] e^{-(\lambda_5 + \lambda_6 + \lambda_7)t}$$

$$3 \left[(1-\rho)(1-e^{-\lambda t}) \right]^2 e^{-\lambda t} \left[1 - (1-\rho_4) \right. \\ \left. \cdot (1-e^{-\lambda_4 t}) \right] \cdot \\ e^{-(\lambda_5 + \lambda_6 + \lambda_7)t}$$

The probability of an extra zero is

$$q_0 = \sum_{i=1}^2 P[X_0(i)]$$

Observe from the set X_0 that transitional errors will be caused by less than three of the input diodes failing through opening. In actuality the probability of a wrong transition for the member $X_0(1)$ in the set X_0 is the joint probability:

$$P(i \text{ th Diode open} \cap \text{"O" on the } i \text{ th input} \cap$$

$$n-1 \text{ diodes working} \cap \text{"1's" on the } n-1 \text{ diodes} \cap \text{no steady-state failures})$$

$$= P(i \text{ th Diode open}) \cdot P(n-1 \text{ Diodes working}) \cdot P(\text{"O" on } i \text{ th input} \cap$$

$$1's \text{ on } n-1 \text{ inputs} \mid i \text{ th Diode open} \cap n-1 \text{ working}) \cdot P(\text{no steady-state failure})$$

The third term in the joint probability expression is the conditional probability expressing the fact that a wrong transition is a function of the information appearing at the gate inputs in any bit time. For all practical purposes this term may be set equal to unity due to the tremendous speed at which information is processed and the resulting short time between occurrence of all possible input states. This same reasoning may be applied to the other member $X_0(2)$.

Note that a NAND gate possesses an asymmetric environment because there are no failure modes that can result in the exclusive classes X_1 or X_{10} .

Thus the reliability of a Transor voting on the output of a network of redundant NAND gates can be defined by equation (10) in part IV.

The following notation was used in this appendix.

- 1) x_i , the event that the i^{th} component is working correctly.
- 2) \bar{x}_i ; the event that the i^{th} component has failed.
- 3) $P(x_i)$; probability of the defined event (1)
- 4) $P(\bar{x}_i) = 1 - P(x_i)$
- 5) x_i^s ; the event that the i^{th} component has shorted
- 6) x_i^o ; the event that the i^{th} component has opened because the probability space of each component is the logical union of

$$x_i \cup (\bar{x}_i \cap x_i^s) \cup (\bar{x}_i \cap x_i^o)$$
- 7) $P(x_i^s)$; the probability of (5)
- 8) $P(x_i^o)$; the probability of (6) = $1 - P(x_i) - P(x_i^s)$
- 9) \bar{x}_i^s ; the event that the i^{th} component has not shorted
- 10) \bar{x}_i^o ; the event that the i^{th} component has not opened
- 11) $P(\bar{x}_i^s)$; the probability of (9)
- 12) $P(\bar{x}_i^o)$; the probability of (10)
- 13) $P(x_i^s | \bar{x}_i)$; the probability of the i^{th} component shorting given that its failed = ρ
- 14) $P(x_i^o | \bar{x}_i)$; the probability of the i^{th} component opening given that its failed. = $1 - \rho$

BIBLIOGRAPHY

- 1) J. von Neuman, "Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components in Automata Studies, "Ed. C. E. Shannon and J. McCarthy, Princeton University Press, 1956.
- 2) W. H. Pierce, "Adaptive Vote-Takers Improve the Use of Redundancy, " Redundancy Techniques for Computing Systems. " Ed. R. H. Wilcox and W. C. Mann, Spartan Books, 1962. July 17, 1961
- 3) "A Survey of Adaptive Components for Use in Failure Free Systems", Special Technical Report No. 1, Nasw-572, Aug. 1963.
- 4) W. C. Mann, "Restorative Processes for Redundant Computing Systems, " Redundancy Techniques for Computing Systems, Ed R. H. Wilcox and W. C. Mann, Spartan Books, 1962.
- 5) A. R. Helland, W. C. Mann, "Failure Effects in Redundant Systems, " Report No. EE-3351, Westinghouse Electronics Division 1963.

Appendix 5

COMPARISON OF DYNAMIC AND THRESHOLD RESTORERS

by

C. G. Masters

R. S. Bray

December 1963

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
I. INTRODUCTION	5-1
II. DESCRIPTION OF DYNAMIC RESTORING CIRCUITS	5-3
A. Review of the Transor Decision Function	5-3
B. Description of the Hamming Distance Restoring Function	5-4
C. Comparison of Transor and the Hamming Distance Restoring Circuit	5-5
III. REVIEW OF THE ANALYTICAL EFFORTS	5-7
A. Signal Processor Assumptions	5-7
B. Classification of Failure Effects	5-8
C. Class Probability Measure	5-10
D. Analytical Models	5-11
1. Multinomial Model for a Dynamic Restoring Circuit	5-11
2. The Transor Model	5-12
3. The Hamming Distance Restoring Circuit Model	5-12
4. The Threshold Restoring Circuit Model	5-13
E. Threshold Parameters as a Bound on Dynamic Parameters	5-14
F. A Comparison of Transor and the Hamming Distance Restoring Circuit	5-16
IV. SIMULATION PROGRAM	5-19
V. DISCUSSION OF RESULTS	5-21
A. Simulation Results	5-21
B. Curves Discussion	5-24
VI. CONCLUSIONS	5-29

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
1.	Block Diagram of the Transor	5-4
2.	Block Diagram of the Hamming Distance Restoring Circuit	5-4
3.	Possible Five-Input Sequences for Two Failures	5-9
4.	Typical Histogram	5-22
5.	Approximation to Reliability Curve	5-22
6.	Transor Order 5 Redundancy	5-23
7.	Comparison of Transor and Hamming Distance	5-25
8.	Comparison of Threshold and Hamming Distance	5-26
9.	Comparison of Order 7 Threshold and Order 5 Hamming Distance	5-28

I. INTRODUCTION

The basic function of a restoring circuit is discussed in Part One of Special Technical Report No. 4 which is contained in Appendix 4 of this report. The Transor is described in that report as a device which is potentially useful for performing the restoring function. Because it is sensitive only to changes in the states of its inputs, a restoring circuit of this type appears to have advantages over the common threshold voter in environments where most failures result in steady state inputs to the restorers. Of course, such a circuit should be inferior to the threshold voter when failures result in transient errors.

The original goal of this study was the determination of the ratio of probability of steady-state errors to probability of transient errors for which any decrease in the ratio will make the use of threshold voter advantageous compared to the Transor. In the process of performing the study, a new dynamic restoring circuit has been developed which has obvious advantages over the Transor for certain input failure pattern conditions. The invention of the Hamming Distance Restoring Circuit caused a shift in the primary goal to include evaluation of both it and the Transor relative to each other, as well as to the threshold voter.

Section II of this report includes a brief review of the Transor and describes the Hamming Distance Restoring Circuit. Section III reviews the analytical techniques which have been used in searching for tools to evaluate the two restorers. Section IV describes the computer simulation program which was used in the evaluation. Sections V and VI contain the results which have been obtained and the conclusions which can be drawn from these results.

II. DESCRIPTION OF DYNAMIC RESTORING CIRCUITS

A. REVIEW OF THE TRANSOR FUNCTION

The Transor is described in detail in Appendix 4. A brief review of the Transor function is given here to ease the discussion of the Hamming Distance function and to facilitate a rough comparison of the salient features of each.

A block diagram of the Transor Restoring Circuit with binary inputs (x_1, x_2, \dots, x_R) is shown in figure 1. The functional relationship between the output Z , the inputs, and the thresholds T_0 and T_1 is expressed in general as

$$Z^{(t)} = f \left[Z^{(t-1)}; (x_1, x_2, \dots, x_R)^t; (x_1, x_2, \dots, x_R)^{(t-1)}; T_0; T_1 \right] \quad (1)$$

The specific function summarized by this relationship may be described as follows. The number of binary "ones" appearing on the Transor inputs during each bit time (t) are summed and compared with the number present during the previous time period $(t-1)$. If the change is positive and greater than a given threshold T_1 then the output Z is forced to a binary "one". If the change is negative and greater in magnitude than a second threshold, T_0 , the output is forced to a binary "zero". If neither threshold is exceeded, the output does not change from its previous state. This operation may be completely specified by the following decision rule statements:

$$\begin{aligned} \sum_{i=0}^R x_i^{(t)} - \sum_{i=0}^R x_i^{(t-1)} &\geq T_1 \longrightarrow Z^{(t)} = 1 \\ \sum_{i=0}^R x_i^{(t)} - \sum_{i=0}^R x_i^{(t-1)} &\leq -T_0 \longrightarrow Z^{(t)} = 0 \\ -T_0 < \sum_{i=0}^R x_i^{(t)} - \sum_{i=0}^R x_i^{(t-1)} &< -T_1 \longrightarrow Z^{(t)} = Z^{(t-1)} \end{aligned}$$

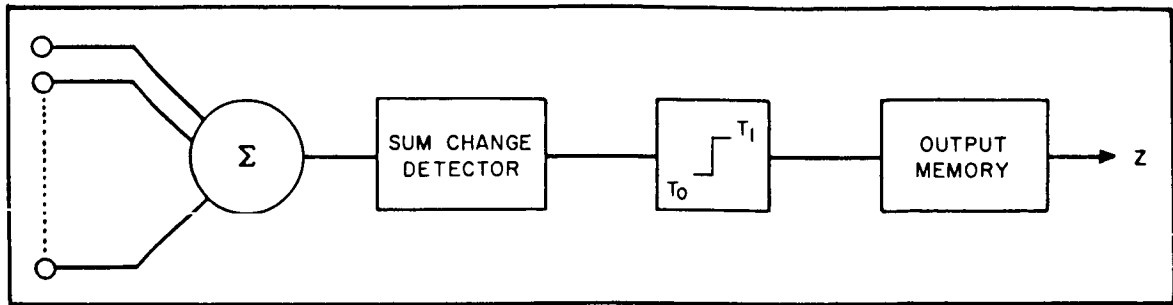


Figure 1. Block Diagram of the Transor

B. DESCRIPTION OF THE HAMMING DISTANCE RESTORING CIRCUIT DECISION FUNCTION

A block diagram for a Hamming Distance Restoring Circuit with binary inputs (x_1, x_2, \dots, x_R) is shown in figure 2. The functional relationship between the output Z , the inputs, and the threshold T can be expressed in a form similar to that of Transor:

$$Z^{(t)} = f \left[Z^{(t-1)}; \begin{matrix} x_1^{(t)} - x_1^{(t-1)}; x_2^{(t)} - x_2^{(t-1)}; \dots \\ x_R^{(t)} - x_R^{(t-1)}; T \end{matrix} \right]$$

Again, this relationship summarizes a rather complicated function. In the same manner as the Transor, the output of the Hamming Distance Restoring Circuit tends to remain in the $Z^{(t-1)}$ state unless the number of state changes on its inputs exceeds some threshold. In the latter case, however, the direction of state changes is not considered and output state change decisions are made without any consideration of the absolute states of the inputs. Thus, the

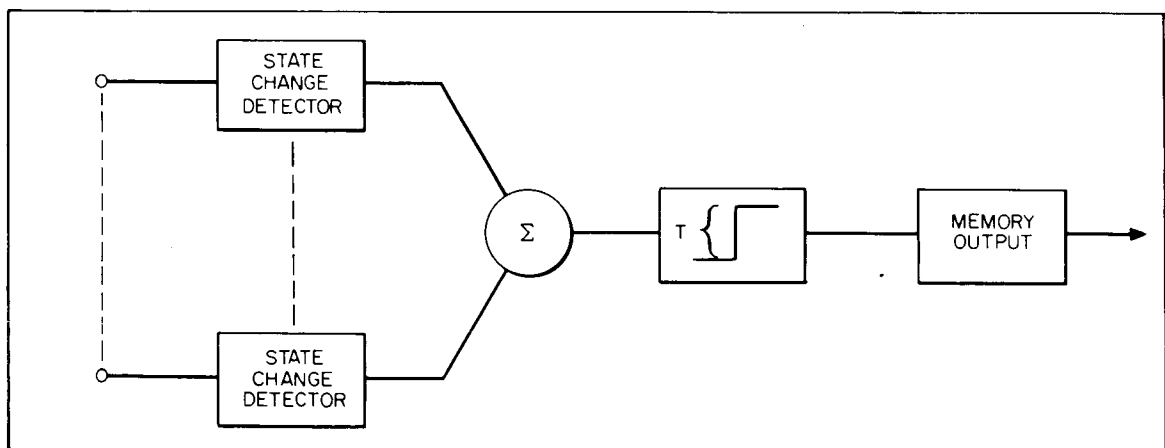


Figure 2. Block Diagram of the Hamming Distance Restoring Circuit

output at time t , $Z^{(t)}$, is always dependent upon $Z^{(t-1)}$ and the Hamming distance between the two input vectors $(x_1, x_2, \dots, x_R)^{(t)}$ and $(x_1, x_2, \dots, x_R)^{(t-1)}$. This relationship is completely specified by the following rule statements*:

$$T > \sum_{i=1}^R |x_i^{(t)} - x_i^{(t-1)}| \longrightarrow Z^{(t)} = Z^{(t-1)}$$

$$T \leq \sum_{i=1}^R |x_i^{(t)} - x_i^{(t-1)}| \longrightarrow Z^{(t)} = \overline{Z^{(t-1)}}$$

C. COMPARISON OF TRANSOR AND THE HAMMING DISTANCE RESTORING CIRCUIT

The outstanding characteristic of the Hamming Distance Restoring Circuit which differentiates it from the Transor is that it ignores information about the absolute state of its inputs. This characteristic can be used to advantage because the input from a signal processor producing both erroneous "ones" and "zeros" cannot cancel the influence of a working processor input as it can in the Transor case. This may be illustrated by considering the following input pattern for two bit times. Suppose that input 3 is failed to a steady state "zero", that inputs 1 and 2 represent the correct information, and that inputs 4 and 5 are producing both extra "ones" and "zeros" at these bit times.

INPUTS	$x_i^{(t-1)}$	$x_i^{(t)}$
1 (correct)	0	1
2 (correct)	0	1
3 (failed)	0	0
4 (incorrect)	1	0
5 (incorrect)	1	0

* The function $\sum_{i=1}^R |x_i^{(t)} - x_i^{(t-1)}|$ is a measure of the difference between vectors $x^{(t)}$

and $x^{(t-1)}$ which applies frequently in information theory. The conception of this measure is credited to R. W. Hamming of Bell Telephone Laboratories.

OUTPUTS	$Z^{(t-1)}$	$Z^{(t)}$
Threshold (majority, T = 3)	0	0
Transor	0	0
Hamming Distance Restorer	0	1

Actually, the states indicated by inputs 4 and 6 need not necessarily occur as a result of component failures. For example, if no provision is made for synchronization, corresponding elements of a redundant binary counter may become permanently out of phase as the result of either noise, or the initially random states due to application of power. For this example, the net change in the number of "ones" is zero, but the total number of state changes is four. It cannot be said from this one example that the Hamming Distance Restorer can always withstand more input failures, but grounds for further consideration have certainly been established.

It should be noted at this point that ignoring the absolute state of the inputs provides the major advantage of the Hamming Distance Restorer but it also a disadvantage. Because the output Z is not directly related to the absolute states of the input, the output state must be set to the correct initial state before operation is begin or it has only a chance, perhaps 50%, of being correct. If it is not initially correct, $Z^{(t)}$ will always be in the state opposite to the correct one. Transor, on the other hand, will converge to the correct value after a small number of bit times because of its dependence on the direction of state changes.

The remaining sections of this report will describe the efforts which have been made to evaluate both Transor and the Hamming Distance Restoring Circuits. These evaluations are referenced to the commonly used threshold voter. The results of the evaluations are discussed in Section V. The conditions under which one of the dynamic restoring circuits might be more powerful than the threshold voter are established.

III. REVIEW OF THE ANALYTICAL EFFORTS

A. SIGNAL PROCESSOR ASSUMPTIONS

To clarify the description of the analysis of the various restoring circuits, it seems advisable to summarize the assumptions which have been made concerning the signal processors which provide inputs to the restoring circuits. Each processor is assumed to be composed of a set of components, all of which must work properly in order for the processor output to be correct. It is assumed that the i -th component of the set has a probability of failure during the differential interval Δ_t which is proportional to the interval length. This probability can be expressed as $\lambda_i \Delta_t$. This implies that the reliability (the probability that the i -th component does not fail during a time interval, t) given by the expression

$$R(t) = e^{-\lambda_i t} \quad (3)$$

Because correct operation of all components is required for correct processor operation and assuming independence of failures between signal processors, the reliability of a processor composed of N components is equal to the product of the component reliabilities.

Therefore:

$$R_s = \prod_{i=1}^N R_i = \prod_{i=1}^N e^{-\lambda_i t} = e^{-\left(\sum_{i=1}^N \lambda_i\right) t} \quad (4)$$

Similarly, if the set of components is partitioned into M subsets and a reliability computed for j -th subset, the processor reliability would be the product of subset reliabilities. Mathematically, this is expressed as

$$R_s = \prod_{j=1}^M R_j \quad (5)$$

and

$$R_j = \prod_{i=1}^{n_j} R_i = e^{-\left(\sum_{i=1}^{n_j} \lambda_i\right) t} \quad (6)$$

where n_j is the number of components in j-th subset and λ_i is the failure rate of the i-th component of the subset. The subsets which the components are partitioned into correspond to the class of processor output errors which failure of the component will cause. The classification of errors is discussed in this section.

If all failure modes of a component caused only errors of one class, the assumption could be made that each component was completely associated with one of the class subsets. In general, this is not true. For example, if the output transistor of a binary signal processor is shorted (emitter to collector), the output would probably become permanently fixed at the "zero" level. If, however, the transistor is open circuited, the output of the processor would probably become permanently fixed at the "one" level. Because subsets are established by classification of output error types: the above transistor cannot be uniquely associated with any subset. To make an association, some artificial method must be used to assign to each subset only that "portion" of a component which will cause that particular class of output error. Although the components cannot be physically divided in the required manner, they can be analytically split by multiplying the total failure rate of the component by the conditional probability of the occurrence of each possible failure mode. This procedure produces a number which can be considered the failure rate of a smaller component or subcomponent whose failure results in only one of the possible classes of output errors.

It should be noted at this point that the failure probabilities of the smaller subcomponents described above are not independent of the operational state of all other similar components, as are the original circuit components. This may be illustrated by referring to the previous example. If the transistor in the example were split into two subcomponents representing the short and open failure modes, and one of the subcomponents had failed, the other component could not also fail. The occurrence of a double failure of subcomponents associated with a single physical component, however, is normally a relatively improbable event in comparison to the other system-failure producing events in associated circuits. For this reason, this dependence effect has been ignored in all the models developed during this study.

B. CLASSIFICATION OF FAILURE EFFECTS

In the initial phase of this study, which is reported in Appendix 4, it was shown that the ability of dynamic restorers to differentiate between inputs working correctly and those failed to a steady state could generate failure modes different from those of threshold decision. There are, specifically, four modes which threaten the operation of dynamic restoring circuits.

- 1) Wrong transitions cancelling correct transitions. (A sufficient number leave a net number of correct signals insufficient to span the set threshold.)

- 2) Wrong transitions occurring while the correct inputs remain the same state (a series of extra "ones" or "zeros"). During this time the nominally correct inputs have lost their voting power so that, if enough wrong transitions occur at one time, they will span the threshold and result in a wrong decision.
- 3) Wrong transitions temporarily simulating steady state failures. Wrong transitions can combine on adjacent bit times in a manner to produce a steady state effect.
- 4) Steady-state failures. Enough steady-state failures would leave insufficient correct signals to span the threshold.

To illustrate, consider figure 3 where state vectors are used to represent the five inputs to Transor. Inputs x_1 and x_2 are assumed to have failed and capable of error. For definiteness all inputs at time (t) may be assumed correct. In the following bit times (proceeding to the right) several failure patterns are possible for each nominally correct input state. The cancellation mode (1) is clearly shown in the sequence (2) \rightarrow (5) where extra "zeros" have appeared at time (t+1). By virtue of the Transor decision rules, an error will be made at (t+2) unless $T_0 = 1$ since the net result of the summation over (t+1) and (t+2) is minus one. Of course, it is also possible for errors to cancel each other as in sequences (3) \rightarrow (4) and (3) \rightarrow (7).

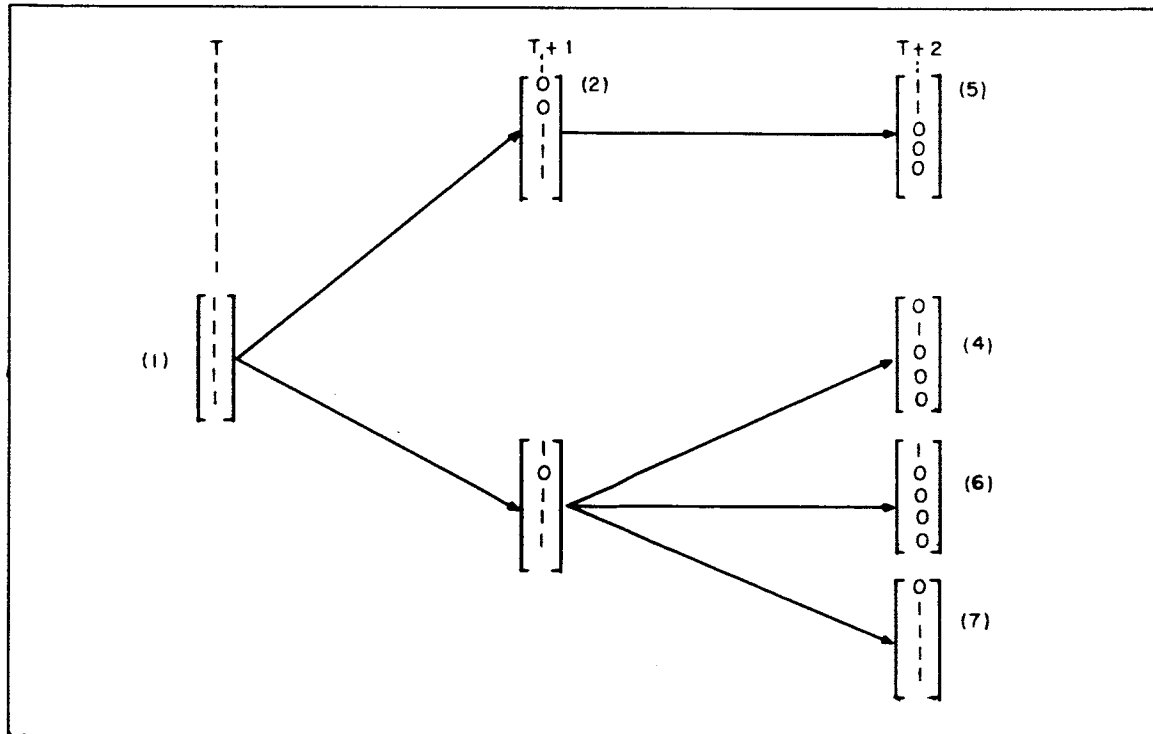


Figure 3. Possible Five-Input Sequences for Two Failures

The second failure mode (2) is shown in sequences (1)→(2) and (1)→(3) and the third mode (3) by sequence (3)→(6). The result is the same in the third mode whether the errors are caused by wrong transitions or steady-state errors.

Any output of a binary signal processor can be classified into one of six mutually exclusive classes over an arbitrary time interval of six mutually exclusive classes over an arbitrary time interval. These are:

- 1) Correct
- 2) Continuous Zero-State
- 3) Continuous One-State
- 4) Extra "ones" but no "zeros"
- 5) Extra "zeros" but no "ones"
- 6) Both extra "ones" and "zeros"

This classification is necessary because the failure modes caused by wrong transitions have no parallel in threshold voter. A realistic comparison cannot be made on the basis of each output simply failing or working. For example, the sixth output mode listed above results in the cancellation effect (1) mentioned earlier. Likewise, output modes (4), (5), and (6) result in the second and third failure modes listed in part A.

C. CLASS PROBABILITY MEASURE

Each of the six mutually exclusive classes must be assigned a separate probability measure. Let these be:

- 1) p : the probability that the output is correct
- 2) $q\gamma_0$: the probability that the output is a continuous "zero"
- 3) $q\gamma_1$: the probability that the output is a continuous "one"
- 4) $q\alpha_0$: the probability that the output generates extra "zeros"
- 5) $q\alpha_1$: the probability that the output generates extra "ones"
- 6) $q\alpha_{10}$: the probability that the output generates both extra "ones" and "zeros" randomly.

The q 's above are related to the reliability of the component subsets through the simple relationship

$$r_j = 1 - q_j \quad \text{where } j = \gamma_0, \gamma_1, \alpha_0, \alpha_1, \alpha_{10} \quad (7)$$

and

$$r = r_{\gamma_0} \cdot r_{\gamma_1} \cdot r_{\alpha_1} \cdot r_{\alpha_0} \cdot r_{\alpha_{10}} \quad (8)$$

Thus, the q 's refer to the probabilities that one or more failures will occur within a particular set of subcomponents and cause the related output error.

D. ANALYTICAL MODELS

In a multiple-line redundant system, it is assumed that each input to a restoring circuit is derived independently, and each input, over an arbitrary time interval, can be defined by one of six mutually exclusive operational classes. A physical system, defined in this manner, suggests a multinomial distribution as its possible analytical model because the R redundant lines can be considered analogous to R repeated trials of an event with more than two possible outcomes.

1. The Multinomial Model for a Dynamic Restoring Circuit

Let the number of outputs failed to a particular mode be represented by a random variable. Specifically, let

γ = the number of outputs failed to the steady state

α_0 = the number of outputs generating extra "zeros"

α_1 = number of outputs generating extra "ones"

α_{10} = number of outputs generating both extra "ones" and "zeros" randomly

Hence, the number of outputs that are continuously correct is

$$R - \alpha_{10} - \alpha_1 - \alpha_0 - \gamma$$

We see that the analytical model for a dynamic voter may be delineated by a subset of points in a four dimensional sample space. These points correspond to possible operating states of the system. Associated with each sample point is a probability defined by the density function

$$\Phi(\alpha_{10}, \alpha_1, \alpha_0, \gamma) = \binom{R}{R - \alpha_{10} - \alpha_1 - \alpha_0 - \gamma, \alpha_{10}, \alpha_1, \alpha_0, \gamma}^* \cdot (P)^{(R - \alpha_{10} - \alpha_1 - \alpha_0 - \gamma)} \cdot (q_{\alpha_{10}})^{\alpha_{10}} \cdot (q_{\alpha_1})^{\alpha_1} \cdot (q_{\alpha_0})^{\alpha_0} \cdot (q_{\gamma})^{\gamma} \quad (9)$$

* The symbol $\binom{N}{x_1, \dots, x_i, \dots, x_m}$ represents the mathematical function $\frac{N!}{\prod_{i=1}^m x_i!}$
 where $\sum_{i=1}^m x_i = N$.

Where

$$p + q_{a_{10}} + q_{a_1} + q_{a_0} + q_{\gamma} = 1 \quad (10)$$

Thus, the reliability of a dynamic restoring circuit will be

$$R(t) = \sum_{\substack{\text{ALL } \Phi \in \Pi}} \Phi(a_{10}, a_1, a_0, \gamma) \quad (11)$$

where Π is the subset of sample points whose outcomes result in a continuously correct decision by the circuit.

2. The Transor Model

For the Transor, membership in the subset Π may be determined by the intersection of the following set of linear inequalities derived from the Transor decision rules.

$$\begin{aligned} a_{10} + a_1 &\leq T_1 - 1 \\ a_{10} + a_0 &\leq T_0 - 1 \\ 2a_{10} + a_1 + a_0 + \gamma &\leq T' \end{aligned}$$

where $\gamma = \gamma_1 + \gamma_0$ and $T' = R - T_0$ or $R - T_1$, whichever is smaller. Thus

$$R_T(t) = \sum_{\substack{\text{ALL } \Phi \text{ SATISFYING} \\ \text{THE DECISION RULES}}} \binom{R - a_{10} - a_1 - a_0 - \gamma}{(p)} \binom{R - a_{10} - a_1 - a_0 - \gamma}{(q_{a_{10}})}^{a_{10}} \binom{R - a_{10} - a_1 - a_0 - \gamma}{(q_{a_1}}^{a_1} \binom{R - a_{10} - a_1 - a_0 - \gamma}{(q_{a_0}}^{a_0} \binom{R - a_{10} - a_1 - a_0 - \gamma}{(q_{\gamma}}^{\gamma} \quad (12)$$

For example, if $R=5$, $T_0 = 2$ and $T_1 = 3$, then

$$R_T(t) = p^5 + 5p^4(1-p) + 10p^3(q_{\gamma})^2 + 10p^3(q_{\alpha})^2 + 20p^3q_{a_1}q_{\gamma} + 20p^3q_{a_0}q_{\gamma} + 20p^3q_{a_1}q_{a_0} \quad (13)$$

3. The Hamming Distance Restoring Circuit Model

The decision rules for the Hamming Distance Restoring Circuit described earlier in the report determine the following set of linear inequalities:

$$\begin{aligned} a_{10} + a_1 &\leq T - 1 \\ a_{10} + a_0 &\leq T - 1 \\ a_{10} + a_1 + a_0 + \gamma &\leq R - T \end{aligned}$$

Removal of the cancellation effect accounts for the absence of the factor of two (2) in the last inequality thus making the Hamming circuit less sensitive to failures causing both extra "one" and "zero" transitions. From these decision rules, the reliability of the circuit can be written as

$$R_H(t) = \sum_{a_{10}=0}^{T-1} \sum_{a_1=0}^{T-1} \sum_{a_0=0}^{T-1} \sum_{\gamma=0}^{R-T-a_{10}-a_1-a_0} \binom{R}{R-a_{10}-a_1-a_0-\gamma, a_{10}, a_1, a_0, \gamma} (P)^{(R-a_{10}-a_1-a_0-\gamma)} (q_{a_{10}})^{a_{10}} \cdot (q_{a_1})^{a_1} (q_{a_0})^{a_0} (q_\gamma)^\gamma \quad (14)$$

For R=5 and T=2

$$R_H(t) = p^5 + 5p^4(1-p) + 10p^3q_\gamma^2 + 20p^3q_{a_1}q_\gamma + 20p^3q_{a_0}q_\gamma + 20p^3q_{a_1}q_{a_0} + 20p^3q_{a_1}q_\gamma + 10p^2q_\gamma^3 + 30p^2q_{a_{10}}q_\gamma^2 + 30p^2q_{a_{10}}q_\gamma^2 + 30p^2q_{a_0}q_\gamma^2 + 60p^2q_{a_1}q_{a_0}q_\gamma \quad (15)$$

4. The Threshold Restoring Circuit Model

In system reliability analysis using majority threshold voters, it is customary to assume that the failure of a majority of inputs, regardless of their mode, will result in a wrong decision. Although this common assumption was used in Special Technical Report No. 4, it is not strictly correct because a threshold voter may tolerate as many as R-1 failed inputs and still function correctly. A more rigorous approach, using the results of section IIB, can be found by letting:

- 1) θ_{10} be a random variable denoting the number of wrong "ones" and "zeros"
- 2) Ψ_1 be a random variable denoting the number of wrong "ones" only
- 3) Ψ_0 be a random variable denoting the number of wrong "zeros" only

Thus, we see that the parameters defined for the threshold voter are related to the dynamic restorer by:

$$\Psi_1 = a_1 + \gamma_1$$

$$\Psi_0 = a_0 + \gamma_0$$

$$\theta_{10} = a_{10} + X$$

where X is a dummy variable which accounts for the case in which a signal processor has experienced two failures causing opposite steady-state errors. Because it is impossible to

say which of the two failure will control the output for a general case, the worst case condition is assumed and in the models both are assumed to exist simultaneously. By virtue of threshold decision rules the subset Π may be defined by

$$\theta_{10} + \Psi_1 \leq T - 1$$

$$\theta_{10} + \Psi_0 \leq R - T$$

The reliability of threshold voter is, then

$$R_{Th}(t) = \sum_{\theta_0=0}^{T''} \sum_{\Psi_1=0}^{T-1-\theta_{10}} \sum_{\Psi_0=0}^{R-T-\theta_{10}} \binom{R-\theta_{10}-\Psi_1-\Psi_0}{\theta_{10}} (p)^{(R-\theta_{10}-\Psi_1-\Psi_0)} (q\theta_{10})^{\theta_{10}} (q\Psi_1)^{\Psi_1} (q\Psi_0)^{\Psi_0} \quad (16)$$

where $T'' = T-1$ or $R-T$ whichever is smaller.

For example, if $R=5$ and $T=3$ we have

$$R_{Th}(t) = p^5 + 5p^4(1-p) + 10p^3(1-p)^2 + 30p^2(q\Psi_1)(q\Psi_0)^2 + 30(p)^2(q\Psi_1)^2(q\Psi_0) + 60p^2(q\theta_{10})(q\Psi_1)(q\Psi_0) + 30p(q\Psi_1)^2(q\Psi_0)^2 \quad (17)$$

E. THRESHOLD PARAMETERS AS A BOUND ON DYNAMIC PARAMETERS

It was shown that the terms in the analytical models corresponded to probability measures associated with specific members of the subset Π within the sample space. Criteria for membership in Π was determined by the intersection of a set of linear inequalities determined from a decision rule.

It will now be shown that a dynamic restoring circuit can now be as effective as the threshold voter when the optimum threshold T for the threshold voter is $(R+1)/2$ and the optimum threshold for a dynamic voter is $\geq (R+1)/2$. It has been shown that when $q\Psi_1 \approx q\Psi_0$ (defined earlier) within a certain range, the optimum threshold for a threshold voter is $(R+1)/2$. The decision for the threshold voter now becomes, using the relations previously described in the threshold restoring circuit model:

$$\theta_{10} + \alpha_1 + \gamma_1 \leq \frac{R-1}{2} \quad (18)$$

$$\theta_{10} + \alpha_0 + \gamma_0 \leq \frac{R-1}{2} \quad (19)$$

Assume also that the ratio of $q_r/(q_{a_{10}} + q_{a_1} + q_{a_0})$ is such that the optimum dynamic restoring circuit threshold is also $(R + 1)/2$; hence, the decision rules for the dynamic circuit becomes

$$a_{10} + a_1 \leq \frac{R-1}{2} \quad (20)$$

$$a_{10} + a_0 \leq \frac{R-1}{2} \quad (21)$$

$$a_{10} + a_1 + a_0 + \gamma_1 + \gamma_0 \leq \frac{R-1}{2} \quad (22)$$

when $\gamma = \gamma_1 + \gamma_0$ and $a_{10} \subset \theta_{10}$. Let all the terms generated by inequalities (18) and (19) form the set Π_{Th} and those by (20), (21), and (22) the set Π_H . The proof consists of simply showing that $\Pi_H \subset \Pi_{Th}$. Clearly each random variable consisted one at a time will form the non-empty sub-sets of the form:

$$\sum_{i=1}^{\frac{R-1}{2}} \binom{R}{R-i} (p)^{R-i} (q_k)^i$$

where $k = \theta_{10}, a_{10}, a_1, a_0, \gamma_1, \gamma_0$. $\Pi_H \subset \Pi_{Th}$ by virtue of the fact that $a_{10} \subset \theta_{10}$. The proof becomes even more obvious when we consider the non-empty subsets formed by combinations of random variables taking two at a time. Choosing one variable from inequality (18) and one from (19) will generate non-empty subsets of the form

$$\sum_{i=1}^{\frac{R-1}{2}} \sum_{j=1}^{\frac{R-1}{2}} \binom{R}{R-i-j, i, j} (p)^{R-i-j} (q_k)^i (q_l)^j \text{ FOR } (k \neq l) \quad (23)$$

where $k, l = \theta_{10}, a_{10}, a_1, a_0, \gamma_1, \gamma_0$. Choosing two terms from (6) will form non-empty subsets of the form

$$\sum_{i+j=2}^{\frac{R-1}{2}} \binom{R}{R-i-j, i, j} (p)^{R-i-j} (q_k)^i (q_l)^j \quad (24)$$

Now $\Pi_H \subset \Pi_{Th}$ because the number of terms generated by (23) is $(\frac{R-1}{2})^2$ and the number in (24) is

$$3 + 4 + \dots + \frac{R+1}{2} = \sum_{M=3}^{\frac{R+1}{2}} M \quad (25)$$

and for all $R \geq 5$

$$\left(\frac{R-1}{2}\right)^2 \geq \sum_{M=1}^{\frac{R-3}{2}} M \quad (26)$$

Likewise, the same reasoning may be applied to combinations of random variables taken three at a time. Thus, it has been shown that if the dynamic restorer is to show superior performance it can only do so when its optimum threshold is reached at values less than $\frac{R+1}{2}$.

F. A COMPARISON OF TRANSOR AND THE HAMMING DISTANCE RESTORING CIRCUIT

In previous discussions, it has been noted that the Transor is controlled by two thresholds as opposed to the single threshold of the Hamming Distance Restoring Circuit. It might be argued that the utility of two thresholds, not necessarily set at the same level, would present an added advantage in a high asymmetrical environment, i. e., one in which either "one" or "zero" errors are more likely. That this is not the case will be shown in the following discussion.

In an earlier Westinghouse report¹ it was shown that in an asymmetrical environment, a great increase in threshold voter performance could be had by using thresholds less than or greater than $(R+1)/2$ according to a criterion developed in that report. Since dynamic restoring circuits cannot distinguish between outputs failed to a continuous "one" and those failed to a continuous "zero", they cannot take advantage of the asymmetry in steady state errors. This leaves for consideration only asymmetrical transitional errors.

The results of the previous section have shown that for a dynamic restoring circuit to show improvement over a threshold voter, the optimum dynamic restoring circuit threshold must be reached at a value less than $(R+1)/2$.

1. P. A. Jensen, "Decision Making in Redundant Systems", Report No. EE-2599, December 1961.

If it is assumed that the optimum value of threshold for the Hamming Distance Restoring Circuit is reached at a value T_{opt} where T_{opt} is less than $(R + 1)/2$ and $R=5$, the following possibilities exist for the Transor thresholds.

$$1) \quad T_0 = T_1 = T_{opt}$$

$$2) \quad T_0 \neq T_1 = T_{opt}$$

$$3) \quad T_1 \neq T_0 = T_{opt}$$

The first case is trivial. If all thresholds are equal, then the Π_T formed from the Transor criteria is clearly a subset of Π_H , i.e., $\Pi_T \subset \Pi_H$ by virtue of the factor $2a_{10}$ in the Transor inequality.

In case (2) T_0 can either be greater or less than T_1 . If $T_0 < T_1$ then $(R - T_1) < (R - T_0)$ and is the controlling factor. But since $(R - T_1) = (R - T_{opt})$ $\Pi_T \subset \Pi_H$. If $T_0 > T_1 = T_{opt}$ for example, $T_0 = T_1 + 1 = T_{opt} + 1$ then $(R - T_0)$ is the controlling factor. But $(R - T_0) = (R - T_{opt} - 1)$ so that, effectively, while the number of terms containing transient probabilities has been increased, the number of terms containing steady-state probabilities has been decreased by the same number and since $q_Y \gg q_{a0}$ the reliability of the Transor will never be as good as that of the Hamming Distance restoring circuit. The same reasoning may be applied to case (3).

IV. SIMULATION PROGRAM

The success of the computer simulation program in evaluating self-repairing systems encouraged the use of a similar program for use as an analytical tool in this phase of the failure free systems study. Such a computer program has been written and has provided a variety of interesting results. Insights into the Transor circuit's most vulnerable areas were gained through this program. One of the results was the development of the Hamming Distance Restoring Circuit. The development of the system failure criteria statements for the program contributed to the development of the general decision rules which have been defined for Transor, Hamming Distance, and Threshold restorers. The program was used to find the ratio of steady-state to transient error probabilities for which the dynamic restoring circuits were at least as effective as the Threshold voter in deriving correct system outputs. Finally, the program provided a check for the analytical models when numerical examples were considered.

Computer simulation programs are commonly used to analyze the performance of deterministic systems which are so large and complex that a mathematical model would be unwieldy or of probabilistic systems which are difficult to model, or when specialized information is desired. The Dynamic Restoring Circuit Evaluator (DRCE) program fell into this last category.

The computer program which has been written for this study retains all of the basic philosophy of the program previously developed for the evaluation of self-repairing systems.* Some portions of the self-repair program were used directly in the DRCE program, but the sections of this latter program which concerned system operational state (i. e., working or failed) are much simpler than those of the self-repair program. These simplifications were possible because of the reduced size and the non-adaptive nature of this simulation problem.

In this simulation program, the range of numbers between zero and unity is divided into intervals, and each interval is assigned to one of the subcomponents of the system. In a system containing (s) subcomponents, the range is divided into (s) intervals each assigned to a different subcomponent. This procedure guarantees that all the numbers in the range are assigned in a manner which uniquely associates every number with only one component and similarly, all components are assigned intervals in the range. By judiciously specifying the lengths of the intervals, random numbers from a population uniformly distributed between zero and unity can be used to simulate naturally occurring random subcomponent failures within the system. To do this, the length of the component interval is made equal to the conditional

*This program is described in Appendix 6.

probability of failure of the subcomponent given that a failure exists somewhere within the system. This probability is given by the expression

$$P_i = \frac{\lambda_i}{R \sum_{i=1}^M \lambda_i}$$

where M is the number of subcomponents in a single processor and R is the order of redundancy (i.e., the number of signal processors in a state). A component failure is simulated by determining a time to failure* and then locating the subcomponent to be designated failed by associating a random number with a particular interval of numbers. Having done this, the type signal processor output error is automatically specified, and the effect of this error on system operation can be found.

As the first step, a system is set up with no initial failures. The above process is begun and continued repetitively until the system under consideration no longer meets one or more operational criteria. At this point, the total system operating time is computed as the sum of the times between component failures. This entire procedure is now repeated many times (usually 100), and data concerning number of failures withstood and system operating times are recorded. From this data various curves are plotted, and system response to various failure patterns is observed.

* The method used to determine the time between each succeeding failure is identical to that used in the self-repairing systems simulation. That method is described on pages 10 and 11 of Appendix 6.

V. DISCUSSION OF RESULTS

A. SIMULATION RESULTS

Before proceeding with a discussion of the results, a brief description of how comparative reliability versus time curves were obtained is required. For each system simulation, the computer print-out includes a number which indicates the total operating time of the system before the occurrence of a critical failure pattern caused loss of system function. These numbers are ordered and split into groups so that a histogram of percent of systems failed versus time can be formed. A typical histogram is shown in figure 4. From this histogram an approximate reliability vs. time curve can be easily constructed by starting a line at unity (100%) on the ordinate and zero (0.0) on the abscissa or time axis and proceeding horizontally to the right until the time corresponding to the first spike on the histogram is reached. At this point the line is dropped vertically by the arithmetic magnitude of the spike, then continued to the right again until the next spike is reached. Continued repetition of this procedure produces a curve such as that shown in figure 5.

The question that immediately arises is "How many system simulations must be run in order for a curve constructed in this manner to be smooth enough to provide a meaningful approximation to the true system reliability curve?" Because the question of "What is smooth enough?" cannot be precisely stated without a series of opinionated assumptions, a simpler, much less rigorous method of evaluation was used. The number of runs was arbitrarily set at 100 and a curve was plotted for a particular Transor voted system. This was compared to a series of points computed from the analytical reliability expression for the same system subject to the same failure rates. The curve and points are shown in figure 6. The correspondence of the curve and the set of points was close enough that the no increase in the number of simulated runs was considered necessary. This relatively low number of runs had the distinct advantage of requiring a computer running time of only about 30 seconds, including compilation time, while producing acceptable results.

One more detail must be pointed out before the curves can be completely understood. The primary interest in the study was the effect which changes the ratio of probabilities of steady state to transient errors. For this reason, the total failure rate of the signal processors was held constant for all simulations. This means that not only the general shape of the reliability curves can be meaningfully compared, but also their locations relative to the time axis. Holding the total failure rate constant in no way restricts the generality of the results because a change in this rate would simply cause a linear shift of the curves along the time axis.

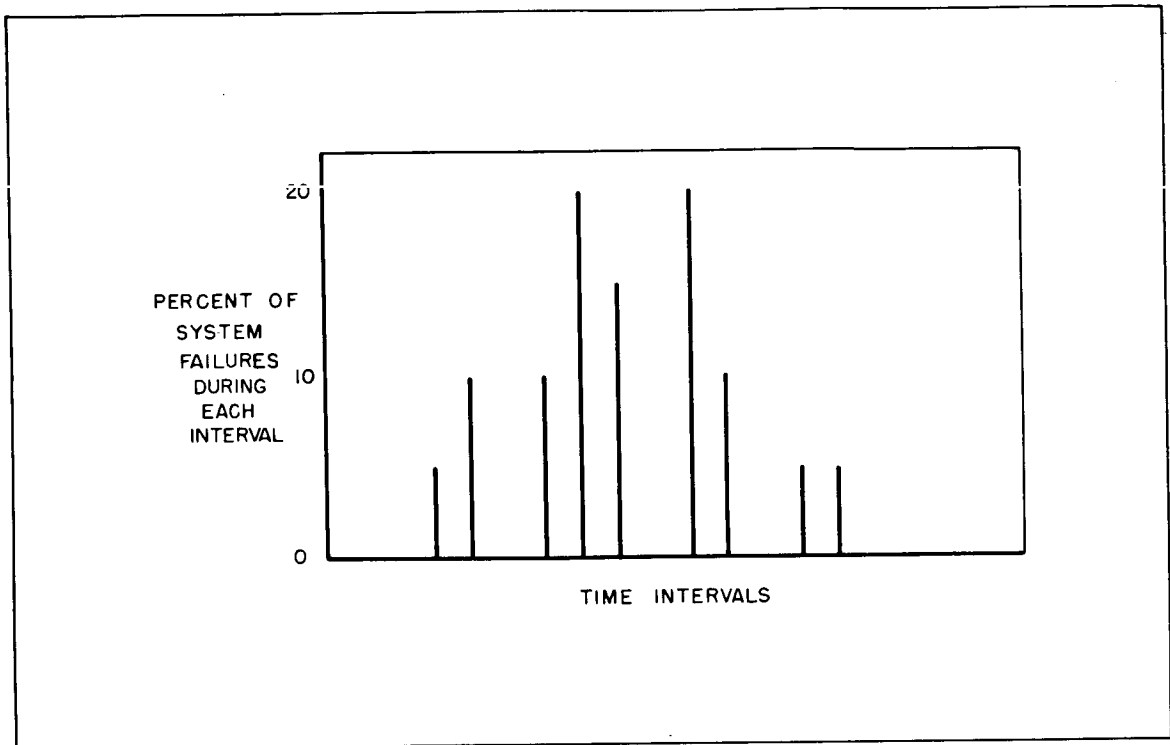


Figure 4. Typical Histogram

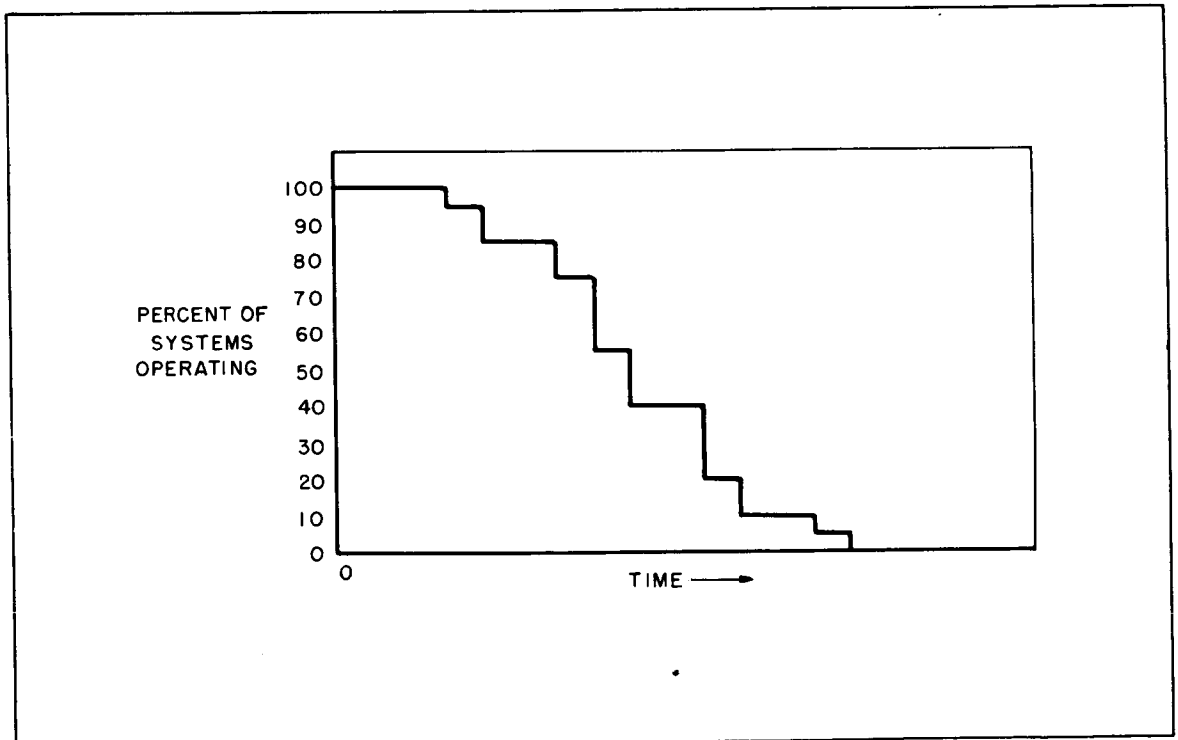


Figure 5. Approximation to Reliability Curve

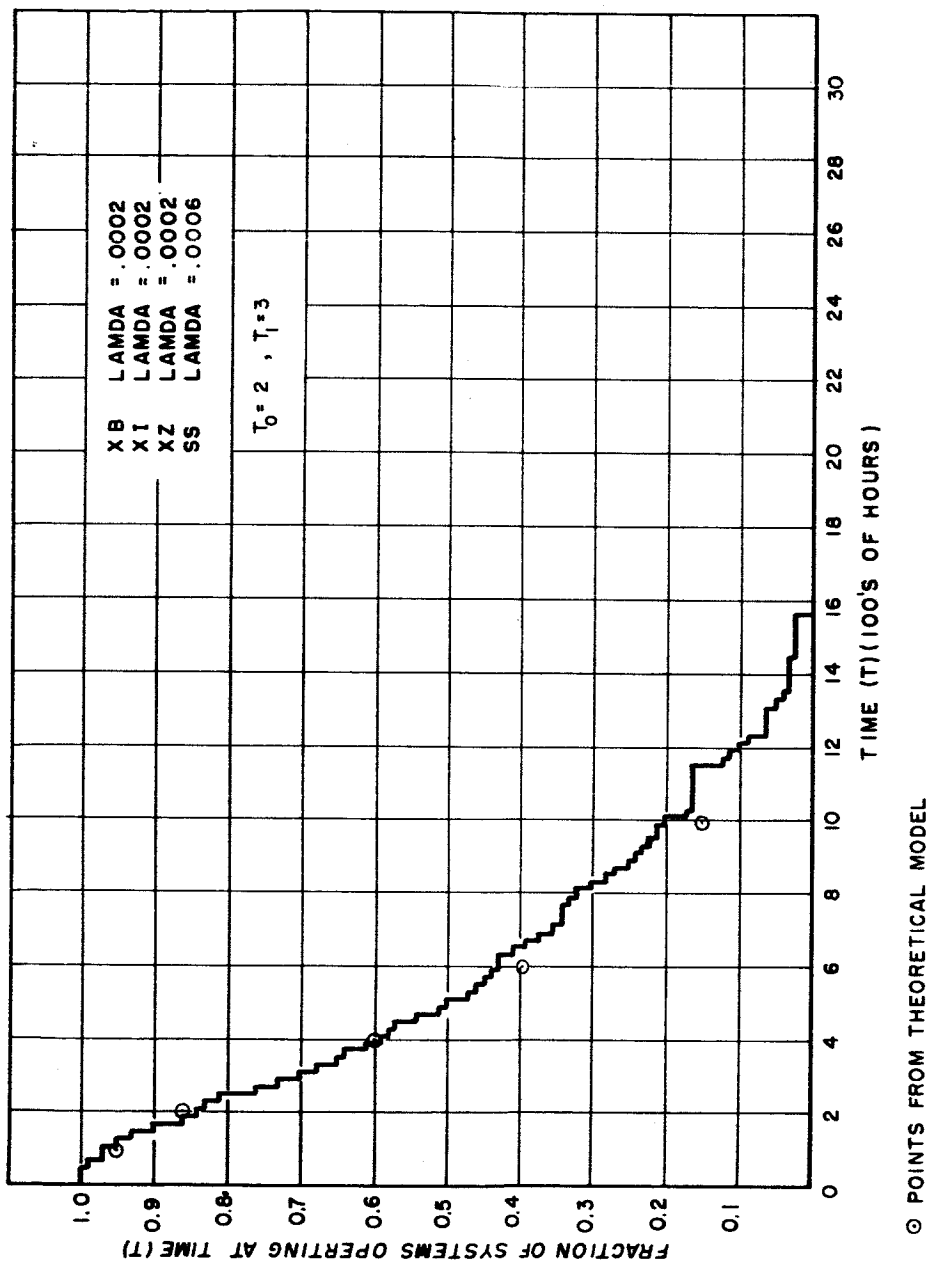


Figure 6. Transor Order 5 Redundancy

B. CURVES DISCUSSION

The first Transor simulations showed that in the region where Transor was competitive to the threshold voter, the optimum T_0 and T_1 were both equal to two for an order five system. The discovery that relationship held even under highly asymmetric failure probability conditions stimulated the development of the Hamming Distance Restoring Circuit. It has since been shown analytically (see Section III) that the Hamming Distance Circuit always dominates the Transor for order five redundancy applications. This result correlates with the simulation comparison for the same configuration, subject to the same failure mode conditions. An example of the simulation results is shown in figure 7.

In comparing the curves for the Hamming Distance Restoring Circuit and those for the threshold voter, it has been found that the latter tends to produce a more reliable output for steady-state to transient error probability ratio below approximately seven to one (7:1) and the Hamming Distance Restoring Circuit slightly more reliable above that ratio. This ratio cannot be exactly determined because certain worst case assumptions have been made in establishing system operational rules for both circuits. These assumptions are slightly more detrimental to one than the other and may not be precisely realistic in either case. This is demonstrated by the combination of points and curves shown in figure 8. In this figure, the Hamming Distance curve appears to be slightly better than the threshold simulation curve in the high reliability region of the curves and worse in the long life region. For this plot of threshold curve, the assumption was made that the first steady-state error to occur in any processor assumed permanent control of the output of the processor and any future transient or steady-state errors in that processor were ignored. The points in that same figure were plotted from a theoretical analysis in which it was assumed that the most detrimental steady-state error which had occurred always controlled the outputs. This worst case assumption does not affect the Hamming Distance curve but it heavily influences the threshold curve. Under this assumption, the Hamming Distance Restoring Circuit clearly dominates over a large section of the curve.

It is interesting to observe the changes which occur in the reliability curves of the restoring circuits as the ratio of steady-state to transient error probabilities is increased. The fact that as this ratio is increased the Hamming Distance curve and the threshold curve get closer together until they cross, indicates that one or both of the curves are shifting in response to the change. The first possibility seems to be the case. The points on the threshold curve tend to remain fixed. (NOTE: a slight shift to the right may be observed. This is caused by a reduction in the pa_{10} as the ratio increases). The Hamming Distance curve

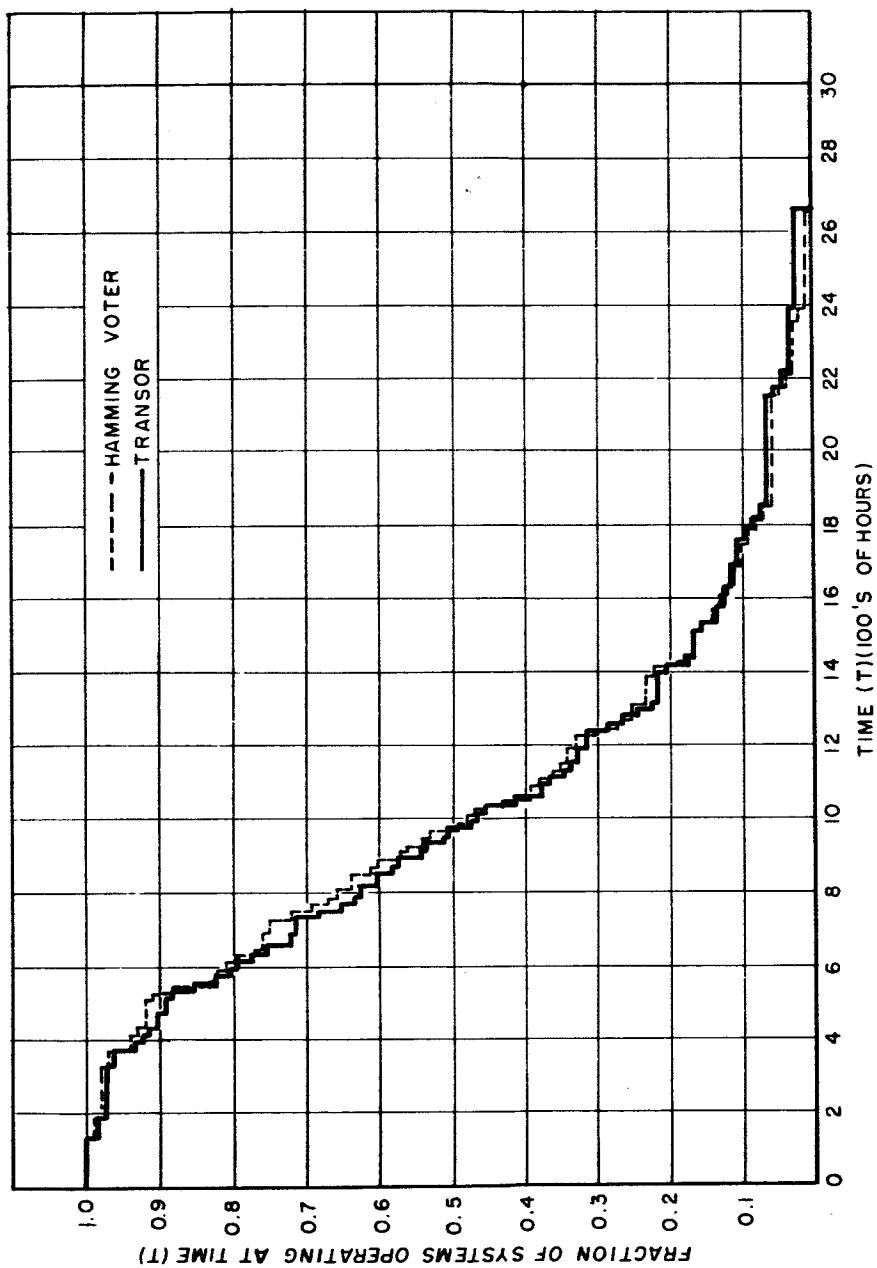


Figure 7. Comparison of Transor and Hamming Distance

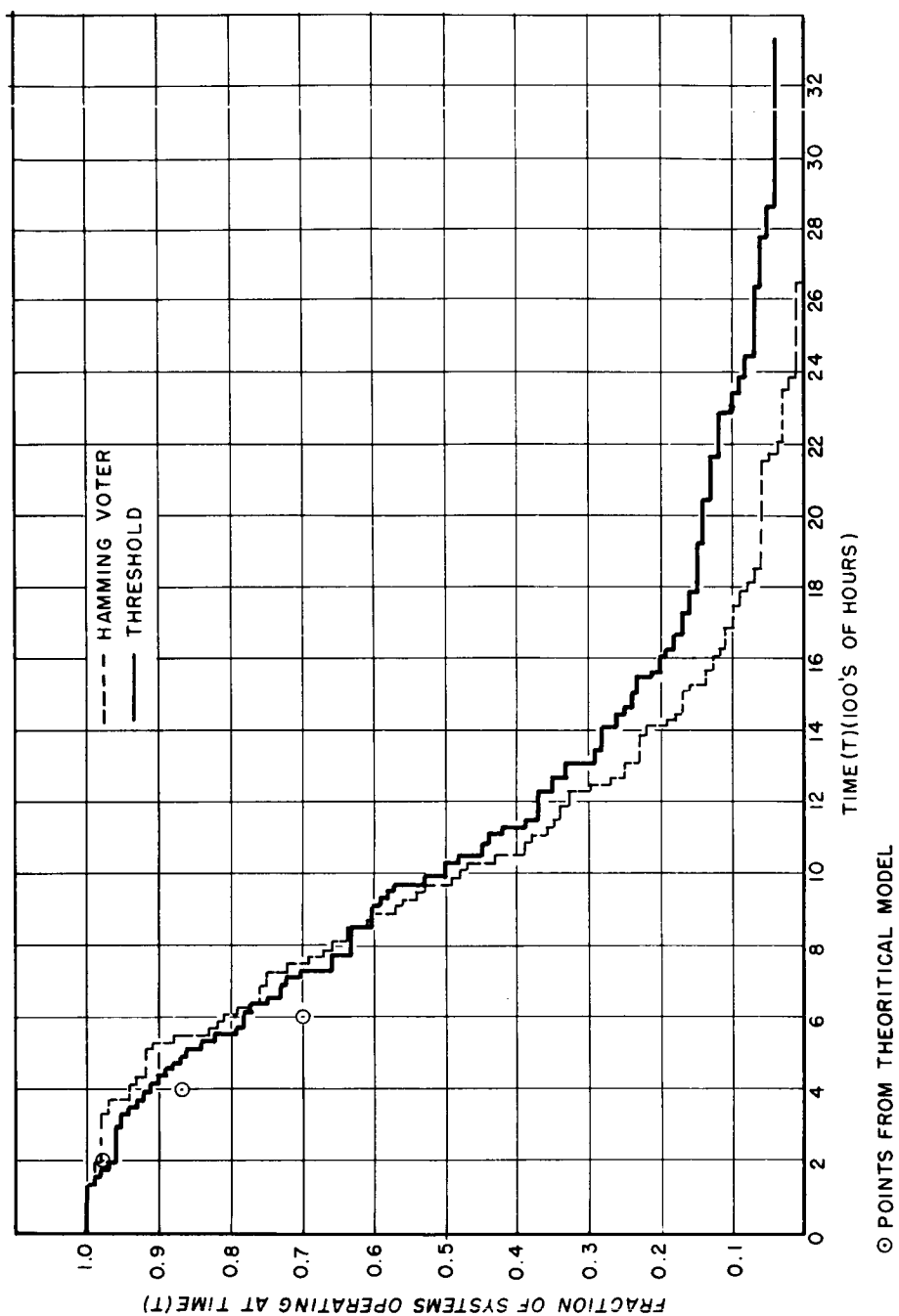


Figure 8. Comparison of Threshold and Hamming Distance

is sensitive to changes in the ratio and shifts rapidly enough to the right to overtake the threshold curve. At approximately the ratio when this occurs, the Hamming Distance curve rapidly becomes less sensitive to changes in the ratio. The ratio continues to be increased, the curve stabilizes and finally begins to slowly fall back to the left, thus indicating that an optimum ratio exists in the region near (7:1). This phenomenon appears to be caused by the discrete nature of the threshold which controls the Hamming Distance decision rules. As the seven to one (7:1) ratio is greatly exceeded, the threshold of the Hamming Distance should be reduced to (1) if additional improvement in the reliability curve is to be expected. This threshold reduction, however, would make the circuit vulnerable to single transient errors. Despite the probable improvement in the overall reliability curve, this sensitivity to single failures is generally considered undesirable. For this reason, no effort was made to simulate systems with this threshold.

In figure 9, a comparison is made between an order five Hamming Distance curve and an order seven threshold curve at a ratio of seven to one (7:1). It can be observed that in the high reliability region, the curves are almost indistinguishable. This implies that under these ratio conditions, an order five Hamming Distance restorer system might be as useful as an order seven threshold voter system. This would allow an obvious saving in redundant equipment.

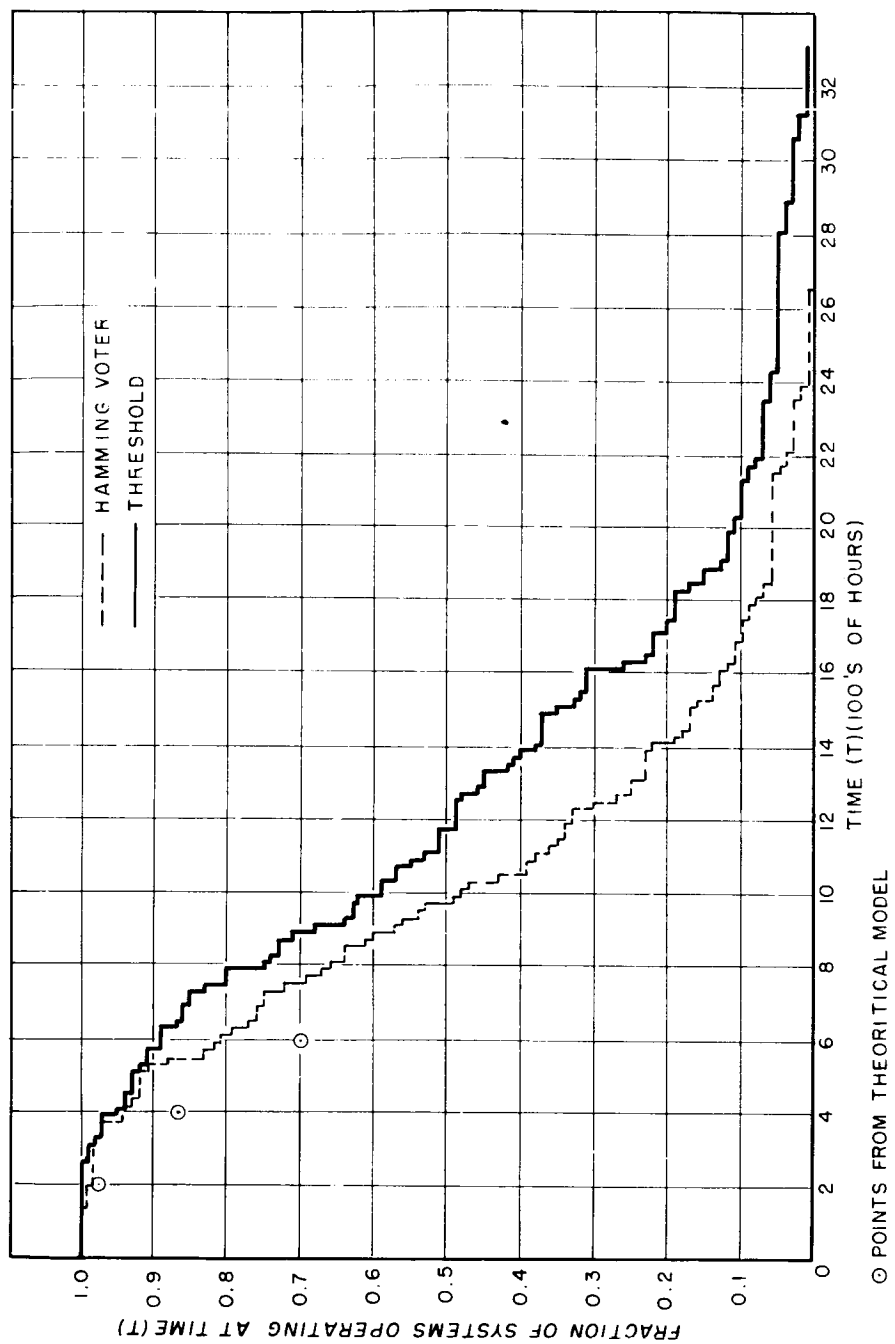


Figure 9. Comparison of Order 7 Threshold and Order 5 Hamming Distance

VI. CONCLUSIONS

From the results obtained by manipulating the analytical reliability expressions for the Transor and Hamming Distance Restoring Circuits, it may be concluded that the output of Hamming Distance Circuit is more reliable than that of the Transor in order five redundant systems. This conclusion holds for any ratio of steady-state to transient error probability or any asymmetry (tendency toward "ones" or "zeros") of error probabilities.

From comparison of the simulation curves, it may be concluded that the threshold circuit is more reliable than either of the dynamic restoring circuits until the ratio of the probability of steady-state errors to the probability of transient error exceeds approximately seven to one. Above this ratio, the dynamic restoring circuit outputs are more reliable. Further comparison reveals that the difference in the reliability curves tend to stabilize or slightly decrease as the ratio becomes much larger than 7:1. The stabilizing effect is more pronounced as the order of redundancy is increased from five to seven.

Finally, it may be concluded that in the short life, high reliability region with approximately a seven to one probability ratio, an order five system using Hamming Distance Restorers may be as reliable as an order seven system using threshold voters.

Appendix 6

SELF REPAIR TECHNIQUES

by

M. R. Cosgrove

C. G. Masters

September 1963

ABSTRACT

This report describes the initial step in the design of an optimal self-repairing system. The report contains a description of the several classes of "repair" strategies under consideration and the computer simulation program which is used to determine the performance of the systems for each strategy.

The computer simulation program determines the performance of a particular strategy by injecting random failures throughout the system and simulating system reaction according to the "repair" pattern of the strategy in question. The program prints out system performance in terms of:

1. total time to failure
2. average time to failure
3. number of failures to system failure
4. number of switches affected.

The results for the two classes of strategies for which curves were drawn show that with the addition of a minimal amount of self-repair capability, the reliability of the system can be substantially increased over that of a comparable system using fixed redundancy alone for failure protection.

TABLE OF CONTENTS

	Page
ABSTRACT.	ii
I. INTRODUCTION	1
II. STRATEGY DESCRIPTION	5
A. Basic Assumptions.	5
B. Basic Strategy Classes Considered to Date	5
III. THE COMPUTER SIMULATION PROGRAM	9
A. The Reason a Simulation Program was Used	9
B. How the Program Works	9
C. Sample Format	12
D. Production Format	13
IV. RESULTS	15
A. Failures Withstood (as percent of system) vs. Spare Mobility . . .	15
B. Reliability vs. Time Curves	17
V. SUMMARY AND CONCLUSIONS.	25
VI. FUTURE STUDIES	27
VII. APPENDIX	29

LIST OF ILLUSTRATIONS

<u>Figure</u>		<u>Page</u>
1	Multiple-line Redundant System.	2
2	Multiple-line Redundant System with Self-Repair Capability	2
3	Probability Distribution of a Component Failure	10
4	Simulation Matrix	12
5	Average Number of Failures Withstood (As Percent of Gamma 1 Systems) Versus Number of Moves per Spare.	16
6	Average Number of Failures withstood (as Percent of Beta Systems) Versus Number of Spares per Block	18
7	Minimum Number of Failures (as Percent of Gamma 1 Systems) Versus Number of Moves per Spare	19
8	Minimum number of Failures (as Percent of Beta Systems) Versus Number of Spares per Block	20
9	Percent of Systems Operating (Beta Class) Versus Time	22
10	Percent of Systems Operating (Gamma Class 1) Versus Time	23

I - INTRODUCTION

In an effort to increase the reliability of complex electronic systems, several methods have been proposed for using "redundant" equipment to provide failure protection within these systems. Two of the most useful types of redundancy techniques are multiple-line, majority voted logic and multiple component grouping schemes. Although both techniques are very effective, a large percentage of the "redundant" equipment is not efficiently used, i.e., the system fails with much of the "redundant" equipment still functioning. This undesirable feature is inherent in systems of this type because random failures do not tend to distribute evenly throughout the system. Instead, they almost invariably tend to group and cause a critical failure pattern to occur in one subsystem area before many failures have occurred in the remainder of the system. The most drastic example of this is the failure of an order three, multiple-line, majority voted system upon the occurrence of two successive failures in the same stage with no other failures in the remaining stages.

Company A has devised a new solution to the failure protection problem which exploits most of the desirable features of the multiple-line, majority-voted schemes, but is not as sensitive to critical failure patterns as the more standard techniques. This solution is in the form of a set of strategies for allowing the reorganization of the systems in response to failure patterns which may develop. The systems which employ these strategies are called self-repairing systems.

The general approach of the self-repair strategies can be described through the use of an example. Figure 1 shows a block diagram of an order three, multiple-line system. Figure 2 shows the same system after some self-repair capability has been added. It is assumed that all blocks in the system are functionally identical such as the multivibrators in a shift register, and are interconnected by switching and voting circuits. If two blocks in the same column fail and the blocks on either side of this column are still operating, the self-repair switching mechanism senses this condition and shifts the required additional working blocks to the failed column. The failed block can now be eliminated or "voted out." This procedure decreases the remaining protection provided the adjacent columns, but it prevents system failure at a critical point and thus extends the life of the system. As additional blocks fail, other blocks are switched into the failed columns. The choice of which block shall be brought in to aid the vulnerable column is determined by the particular strategy in use.

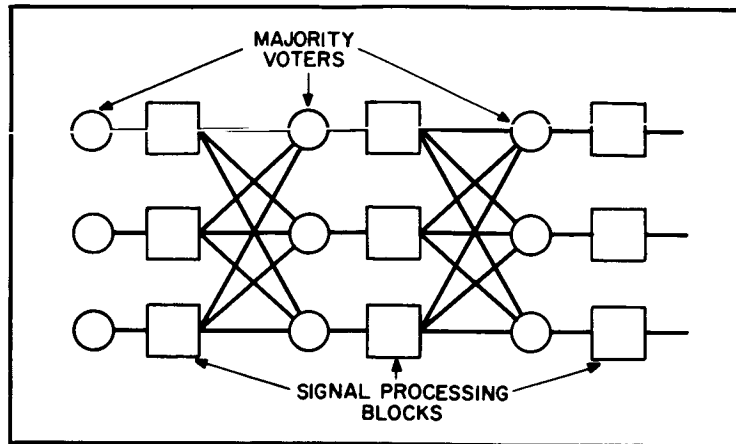


Figure 1. Multiple-line Redundant System

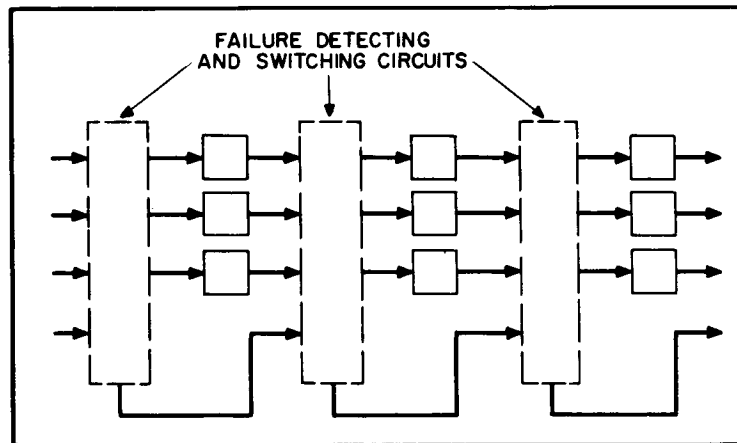


Figure 2. Multiple-line Redundant System with Self-repair Capability

The unique feature of these strategies is that the switching circuitry can be completely distributed rather than "lumped" into a central controller. As a result, most failures in the switching circuitry are equivalent to signal processor (block) failures and are eliminated in the normal manner. This means that individual failures in the switching circuitry do not cause the loss of the entire self-repair capability.

Before a "hardware" design of self-repairing systems can begin, the full range of feasible switching strategies must be examined, and from these an optimum strategy or set of near optimum strategies must be selected. The majority of this report is concerned with

a description of some of the more promising strategies and with the computer program which is being used to simulate the failure response of systems which employ these strategies.

There are a great number of possible strategies which may be investigated, many of which are quite similar to one another. The strategies being considered are arranged in groups called classes, the individual members of which are special cases of the general class. This allows the investigation and programming of a few classes of strategies rather than many individual strategies. This facilitates comparison of strategies within a class as well as adding a certain degree of generality to the analysis.

Before proceeding to the description of specific strategies or classes of strategies, the properties a self-repairing system should have must be noted and the basic assumptions stated. A short list of the general desirable properties is compiled below.

- a. Self-repairing systems should be more reliable than ordinary redundant systems of identical function capability and cost.
- b. The switching strategy used should make optimum use of the redundant function blocks for a fixed amount of switching complexity.
- c. Instantaneous failure masking must be provided for system applications which cannot withstand a temporary loss of data. An example of this is the key-stream generator used in secure communication channels.
- d. The strategy must be suitable for implementation by a distributed (non-centralized) switching network.

II - STRATEGY DESCRIPTION

A. BASIC ASSUMPTIONS

Almost all large computing and control systems are formed by interconnecting a relatively small number of different types of basic circuit blocks. As a result, the components of these systems can be split up into homogeneous groups of functionally similar or identical blocks. It is assumed, therefore, that such groups can be formed and that self-repair strategies can be applied within each group. Note: The members of any group are not required to be physically or functionally adjacent but may be located in scattered sections of the overall system.

It is also assumed that at least two blocks must be performing the same nominal function before a failure can be detected, and at least two correctly operating blocks must be performing the same function before a third (failed) block can be eliminated from this function.

If at least three blocks are performing a function and one of them fails, the elimination process is assumed to be instantaneous, and the failure is assumed to be completely masked. If, however, only two blocks are performing the function and one fails, a third block must be switched to that location to eliminate the failure. This process is not assumed to be instantaneous and errors appear in the system temporarily. As a result, systems using the basic order-three redundancy with self-repair (as will be described in the Beta and Gamma Class strategies of this report) must be capable of withstanding temporary data loss without mission failure. If this assumption is not true, a higher order of redundancy must be used as in the Alpha class strategies or higher-order versions of the Beta and Gamma classes.

If, because of particular failure and response patterns, single blocks are left to perform particular functions it is assumed that the system continues to operate with one or more stages existing in the non-redundant state either until one of these blocks fails or until another critical failure pattern occurs elsewhere in the system.

Finally, it is assumed that a stage shown pictorially at one end of a system is, in reality, adjacent to the opposite end and enjoys the same repair facilities as stages shown in the center of the system.

B. BASIC STRATEGY CLASSES CONSIDERED TO DATE

The following few paragraphs will indicate the general principles of each of the three strategy classes which have been simulated thus far. Detailed examples of each class are shown in the Appendix, and the reader will probably need to refer to these for detailed consideration of the following descriptions.

1. Alpha (α) Class

Systems employing the α class strategies are basically multiple-line redundant (usually order three) systems which are equipped with sets of spares. These spares are additional function blocks which can be automatically used to replace failed blocks. In general, spares can not economically be given enough mobility to allow a single spare to be capable of replacing each operational block in the entire system. Instead, individual spares are usually given restricted capability and may replace only blocks in a single row* or portion of a row. A large number of strategies, each belonging to the (α) class, can be generated by varying (a) the total number of spares available for a fixed system size, (b) the mobility of each spare (c) the pattern in which the spares' repair capabilities overlap.

If it is assumed that spares will immediately replace failed blocks regardless of whether it is the first failure in a function column or not, complete failure masking is achieved. The threshold vote technique will continue to absorb failures after the spares complement is exhausted until a majority of unrepairable failures have occurred at a particular function. At this point the system will fail since both the self repair capability and the network redundancy have been exhausted.

2. Beta (β) Class

Beta Class strategies do not utilize inactive spare blocks as does Class α . With no failures, the system operates as an ordinary multiple-line redundant system. When a critical failure i.e., one which would cause failure of a multiple-line redundant system, occurs, the failed block is removed from the system and replaced by a properly functioning block from an immediately adjacent function. The individual strategies in this class differ from one another primarily in the number of spares which they can draw from the rest of the system.

Because failures are replaced by function blocks only from the adjacent functions there is a smaller amount of switching circuitry involved with Class β than with other classes of self-repair strategies. This advantage is partially offset, however, by the one drawback inherent in this class of strategies. That is these systems are more vulnerable to failures which are grouped in one area of the system than are the more flexible strategies.

The three strategies of this type which have been simulated are described in the Appendix. These particular strategies do not usually allow blocks to move a second time after an initial repair has been made. This restriction has been made for a variety of reasons, but other strategies are being considered which will release this restriction. In addition, strategies having increased spare mobility will be considered in future studies.

* For example the top line or row of signal processor in Figure 1.

3. Gamma (γ) Class

The Gamma (γ) Class of self-repair strategies contains much more variety than either Class α or Class β . The class is characterized by a shifting of the spare blocks in one direction to alleviate the critical condition caused by the failed function blocks. Unlike the strategies of Class β , it is possible for a spare to move several times in response to failures. When a critical failure occurs, one of the function blocks adjacent to the failure will replace it, leaving a void. This void, if it creates a vulnerable situation i. e., one block per function stage, will be filled by the function block immediately adjacent to it in the opposite direction from the original failure. The next failure to occur in the same stage as the original failure causes another shift of the function block now adjacent to the failure. This may be a function which has already shifted in response to a failure. As long as spares are available, they will continue to shift laterally to replace failed blocks or to fill voids.

Since the spare function blocks are allowed much more mobility in this class of strategies, more failures can be corrected. However, the amount of switching circuitry necessary to implement the strategies is a monotonically non-decreasing function of the mobility of the spares. This creates problems of implementation which limit the usefulness of high spares mobility.

The individual members of Class γ strategies differ primarily in amount of mobility allowed to the function blocks. This, in turn, affects the failure absorption capabilities of the strategies. Again, the individual strategies are described in more detail in the Appendix.

III. THE COMPUTER SIMULATION PROGRAM

A. THE REASON A SIMULATION PROGRAM WAS USED

Although the reorganization features of self-repairing systems improve the failure absorption capability of redundant networks, these features drastically affect the analytical reliability expressions developed for multiple-line, majority-voted systems. Not only does a slight amount of reorganization capability greatly complicate the expressions, but each modification of each strategy class appears to require a different solution. Extensive efforts to model some of the simpler self-repairing systems have been unsuccessful. Because of this, efforts to write exact reliability expressions have been dropped, and a general computer simulation program has been written to facilitate a Monte Carlo approach to the reliability analysis. This program can be used to simulate a broad range of strategies, and it provides data about the actual switching patterns which tend to occur in a system. This latter information could not be easily determined from reliability expressions even if they were available. A plot of reliability versus time can be obtained directly from the program results with no more additional input information than would be required by calculations made using analytical expressions.

B. HOW THE PROGRAM WORKS

1. The General Program Philosophy

A redundant system of the desired order of redundancy and number of functions is set up in matrix form. The strategy class is then selected from a group of sub-programs and input data which specifies the particular strategy to be tested is read in. Through the use of a series of random numbers, individual blocks are designated as failed, and the switching strategy responds to each failure until the system fails to pass the operational criteria. A second series of exponentially distributed random numbers determines the time between each simulated failure, and the sum of these is the time to system failure. Once the system fails, the pertinent data is recorded, and the computer resets and begins to generate two new sets of random numbers. Continued repetition of this process provides the compilation of data mentioned in part A of this section. The following paragraphs indicate specifically how the various portions of the program work and the form of the print out.

2. The Failure Selection Program

A simple procedure for randomly selecting the failed function blocks has been set up. Each block is assumed to have an exponentially decaying reliability = $e^{-\lambda t}$ where λ is a constant failure rate. It has been shown that the conditional probability that a failure

has occurred in the i^{th} block given that a failure has occurred in the system is equal to the

$$\text{constant, } \frac{\lambda_i}{\sum_{i=1}^N \lambda_i}.$$

If the interval between zero and one is split into N subintervals, each proportional to the associated conditional probability, a set of random numbers uniformly distributed between zero and one can be used to determine which blocks fail with correct conditional probability of picking any one box. In this particular computer program, the random number specifies the block to be failed. The system then responds to eliminate the failed block. If the response is possible, i. e., a spare block is available to make the repair, a new random number is chosen and the procedure repeats. If no spare is available, the system is judged as failed.

3. Time Determination

For each of the simulated failed blocks selected above, a time to failure for the block is also determined. A. M. Mood¹ has shown that random numbers taken from the uniform distribution can be transformed into any desired continuous distribution by letting

$$f(y) = 1 \quad 0 < Y < 1$$

$$y = G(x)$$

Where $G(x)$ is the cumulative distribution of x .

This relationship is shown graphically in figure 3.

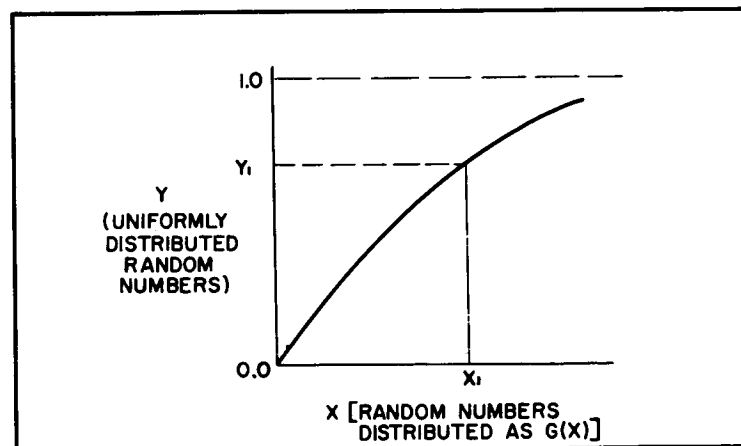


Figure 3. Probability Distribution of a Component Failure

¹Mood, A. M. - Introduction to the Theory of Statistics McGraw Hill Book Co., Inc. 1950

Y is a single valued function of x and vice versa. For each Y chosen from a uniform distribution, a unique value of x is determined.

The G (x) function which is of particular interest here is $G(t) = 1 - R(t) = 1 - e^{-\lambda t}$. This is the distribution function associated with the probability that the first failure has occurred within a system. This curve is shown in figure 3.

For the first function block failure, a random number is chosen from a uniform population and transformed to a corresponding number from the exponential distribution. This latter number is the time from system start to the first failure. To calculate the time to the second failure, the λ associated with the first failed block should be subtracted from the $\Sigma \lambda$'s and the procedure repeated. The new number thus obtained would be the time from the occurrence of the first failure to the occurrence of the second failure. When the system fails, the sum of these individual failure times will determine the total system operating time.

In the present program, the above procedure is slightly modified to make computations easier. Instead of decreasing the $\Sigma \lambda$'s after each failure, this sum is left the same and blocks are allowed to fail more than once. When a block fails for the second time no action is taken other than to add the time to this failure to the system operating time. This modified procedure would not be acceptable if the times between subsystem failures were of interest, but since total system operating time is the only factor to be considered, the results are almost identical to these which would be obtained in the more straightforward approach.

4. The System Reactions

It is obvious that many specific reactions are different for different strategies, but the general manner in which the program performs the various shifts and the type "bookkeeping" involved can be briefly described. Figure 4 schematically illustrates the form in which computer "views" the system to be simulated. The height of the "basic array" is set by the original order of redundancy, the width by the number of stages, and the depth by the number of data words associated with each block. The "failed block array" is a two-dimensional array into which the data words for failed blocks are shifted as the failures occur. The only indication to the computer that a block has failed is the shifting of these data words into this latter array.

When a set of data words is moved into this array, the computer examines the remainder of the system and makes any necessary response. This is done by shifting the data words associated with the appropriate spare blocks from their original locations into the locations specified by the particular switching strategy being considered.

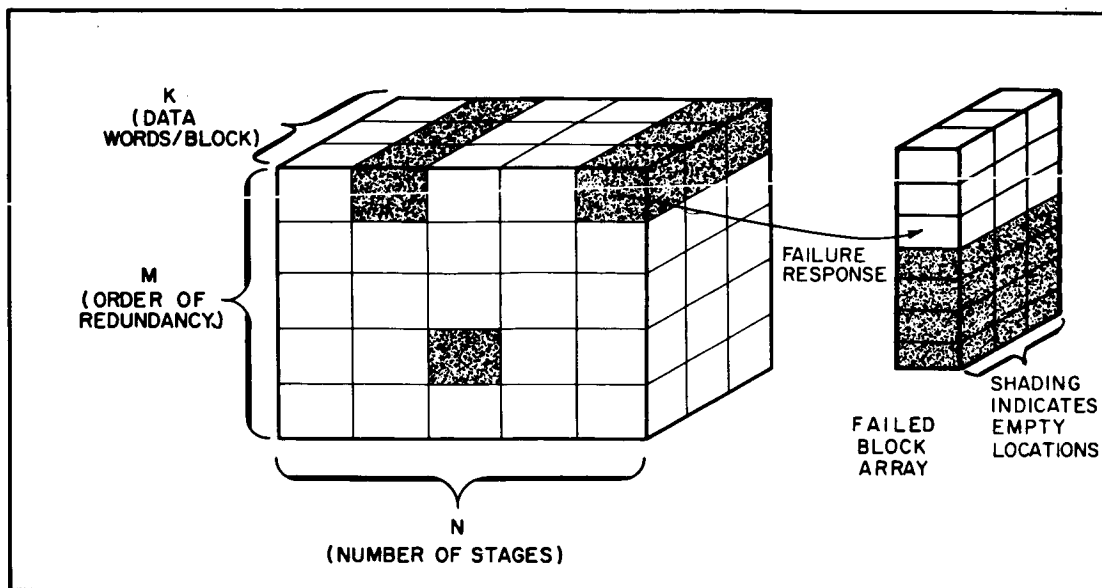


Figure 4. Simulation Matrix

C. SAMPLE FORMAT

A check must be made to determine whether the computer simulation program is operating correctly, i. e., selecting the correct function block for failure according to the random number set, responding properly to failures according to the particular strategy, and failing at the proper time and under the proper conditions. In order to accomplish this, a sample format has been developed. This sample format prints out the following information:

1. * The function block designations and the random number range which describes failure of the block.
2. * A list of failures which occur with all the information associated with the failure such as:
 - a. The random number which was selected
 - b. The location of the failed block
 - c. The amount of time from the previous failure to the time of failure of the block in question
 - d. The cumulative time from the beginning of system operation.
3. The average time between failures.

* This information is printed out for each failure until the system fails.

When a critical failure of a function block occurs, an operating spare is switched into the vacant position by assigning random number limits of the spare block to the failure location. This permits checking of the switching pattern to determine if the simulation program is working, since an incorrect switching operation will place the random number limit designation in the wrong position. This event can be detected when the incorrectly switched function block fails and the position specified by the random number does not correspond to that printed out in the sample format.

To check a strategy, several runs are made using different random number sequences. The sample format prints out all the above information for each case. From this information a determination can be made as to whether the simulation is following the rules for the particular strategy.

In addition to performing the function of checking the simulation program, the sample format provides another valuable service. By observing the vicissitudes of the system with respect to the switching patterns which develop, information can be gained about changes in the strategy which might profitably be used to implement more efficient system operation or more economical switching circuitry implementation. This is the manner in which Class γ_2 was derived from class γ_1 .

D. PRODUCTION FORMAT

A typical production run of the computer program simulates system operation for one hundred randomly selected failure patterns. Up to the present time, all runs have included one hundred patterns simply because relatively good estimates of the average system parameters such as total time to fail, number of failures withstood, etc. are obtained without requiring excessive amounts of computer time.

The production format directly provides the following information for each of the one hundred cases:

1. Average time between function block failures
2. Total time to system failure
3. Total number of function block failures before each system failure
(including multiple failures of the same block)
4. Net number of failed function blocks at time of system failure
5. Total number of switching moves experienced by each system
6. Total number of moves made by each spare function block.

In addition to printing out columns of numbers covering the first five items on the list above, most of the data is compiled into bar graphs. Each of these graphs reflects the

performance of the set of one hundred runs with respect to a particular parameter. On the graphs, either discrete points (e.g. net number of failures) or interval terminal points (for continuous parameters such as time) are plotted on the abscissa. The height of the bar above each point or interval shows the number of spares or system simulations which are described by these positions on the abscissa. The program includes a normalization routine for each graph which is used to compute the average, the variance and the standard deviation associated with each graph.

IV. RESULTS

The strategies discussed here (and any new ones which may be invented) must be compared and contrasted to determine their usefulness in increasing the reliability of electronic systems. The primary goal of this comparison is the determination of which strategy provides the greatest net increase in system reliability. Because it appears that the switching circuitry associated with spare blocks increases as the mobility of these blocks increases and because the failure protection effectiveness of added flexibility is non-linear, it cannot be simply assumed that the best strategy is the one with the greatest spare block mobility.

The best way to compare these strategies would be to completely design functionally identical systems using each strategy; get the best available estimates of the failure rates of all the parts; feed this into the computer program and, in the manner described below, plot the reliability versus time curves. The comparison would merely require that one directly observe which strategy has the highest reliability curve. This approach would require a detailed system design for all strategies. To avoid wasting time on strategies which can be shown to be inferior to others with much less detailed input data, several less exact comparisons can be made. These comparisons, which are described below, are the ones which are being made at this point in the study.

A. FAILURES WITHSTOOD (AS PERCENT OF SYSTEM) vs. SPARE MOBILITY

An important consideration in the comparison of systems is the number of failures which can be withstood without system failure. In order to compare strategies with one another where the variable is the number of moves allowed per spare, the number of failures withstood is an important and meaningful criterion. To further compare systems of different sizes on a common base the curves plotted for these systems are expressed in terms of average percent of total system failed versus spare mobility. In figure 5 curves are plotted for three systems of different sizes, 24, 48 and 96 stages employing strategy γ_1 . They are plots of average percent of failures versus number of moves per spare.

These curves provide very useful and interesting results. They are characterized by a sharp rise, a knee and a rapid leveling off. The knee occurs at a small number of moves per spare compared to complete (total system) spare mobility. According to this graph, a great increase in number of failures withstood by a system is effected by increasing spares' mobility up to a point. The increase, then, is diminished and a point is reached beyond which little or no increase in number of failures withstood accompanies an increase in mobility. The characteristic exhibited by these curves illustrates that great increases can be attained in system performance by the introduction of self-repair Class γ_1 with

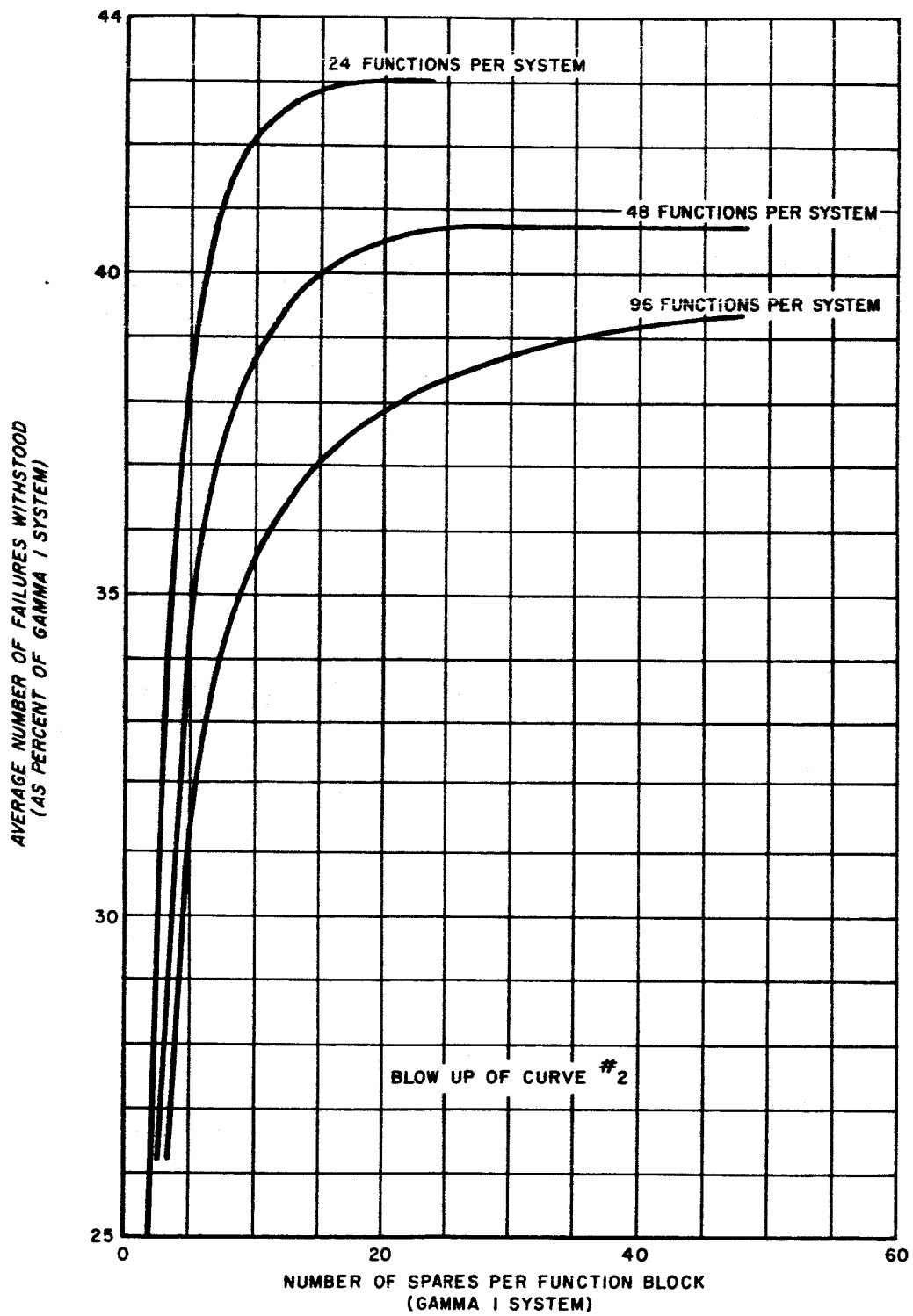


Figure 5. Average Number of Failures Withstood (as Percent of Gamma 1 Systems) Versus Number of Moves Per Spare

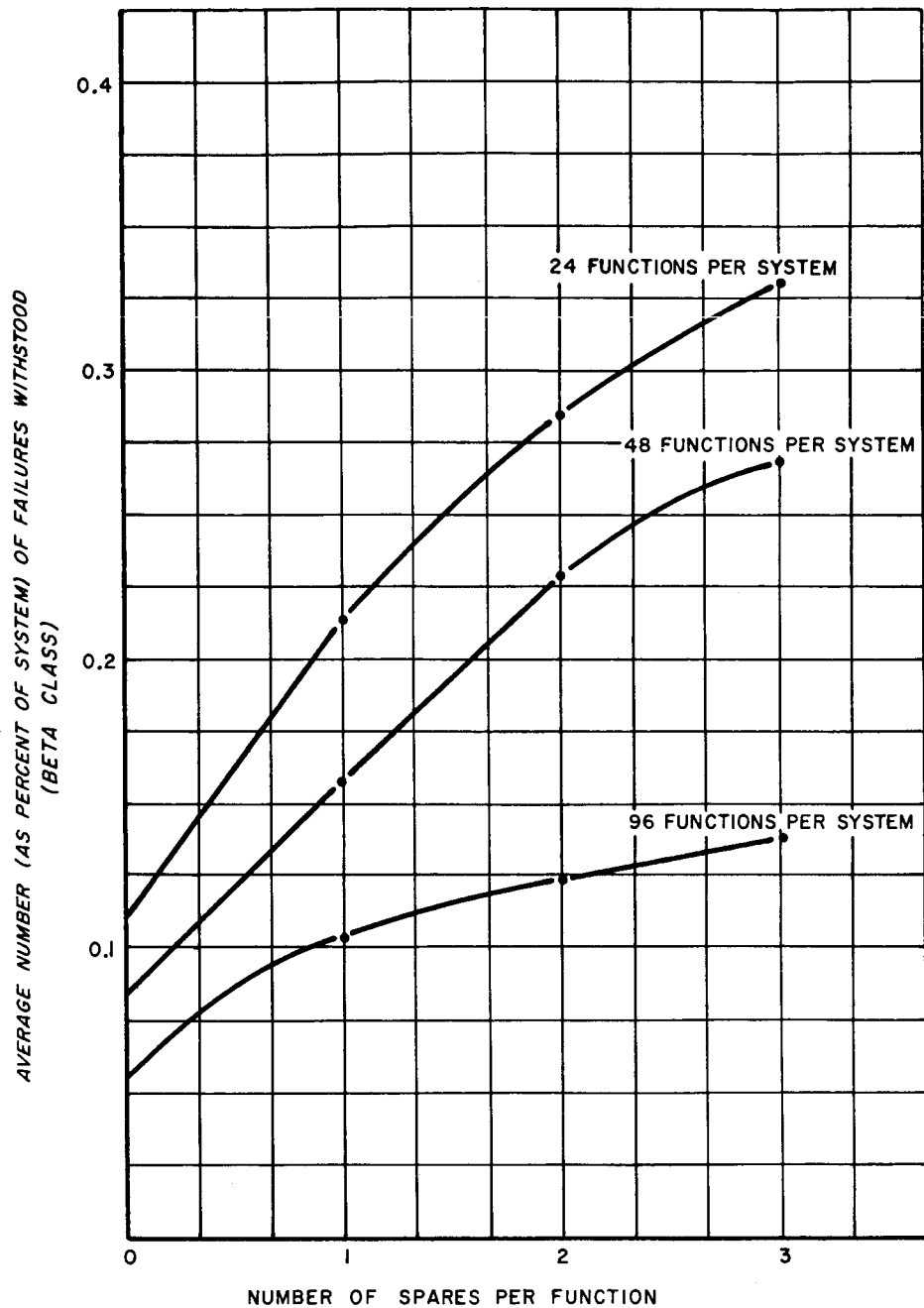


Figure 6. Average Number of Failures Withstood (as Percent of Beta Systems) Versus Number of Spares per Block

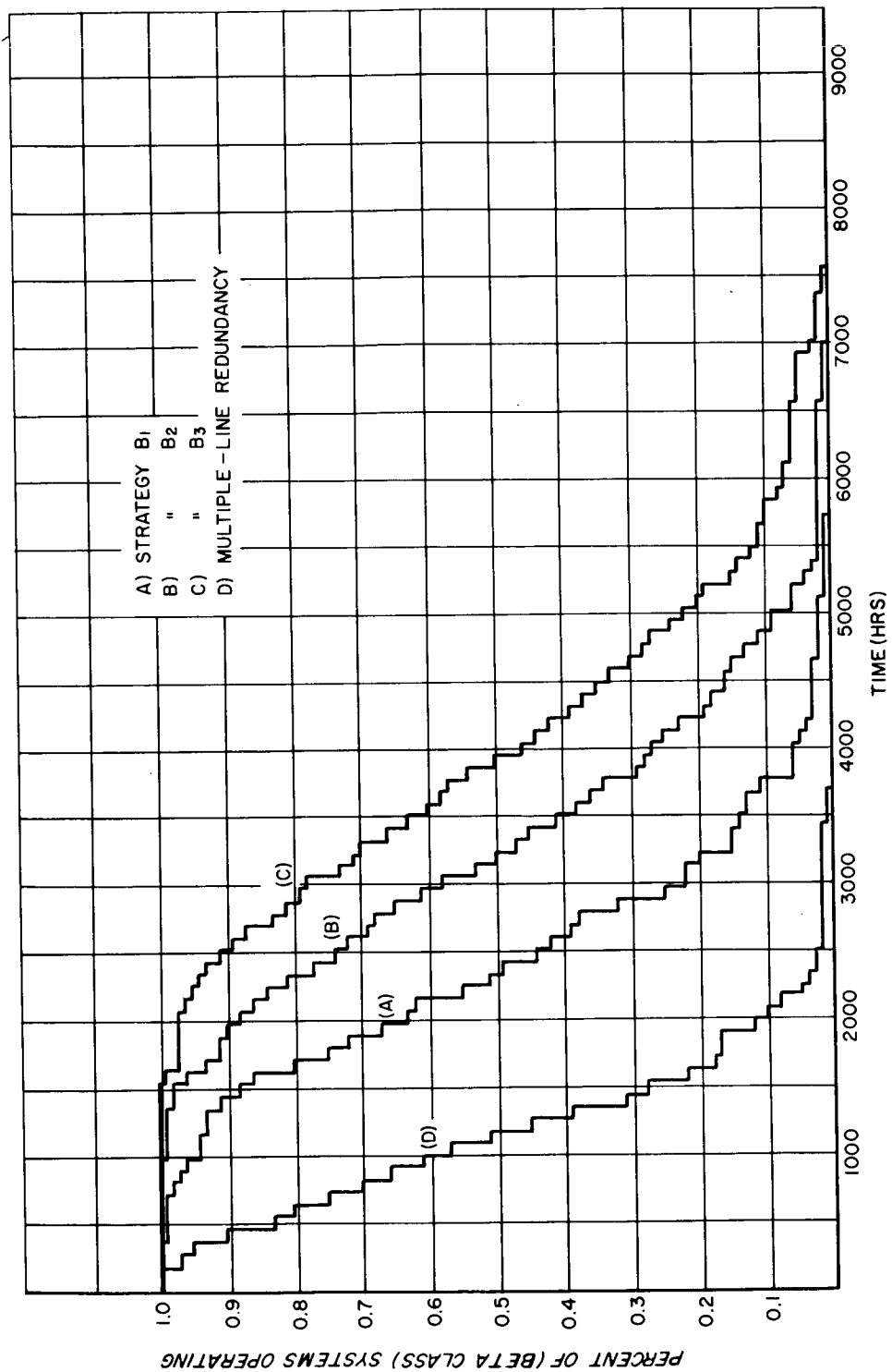


Figure 9. Percent of Systems Operating (Beta Class) Versus Time

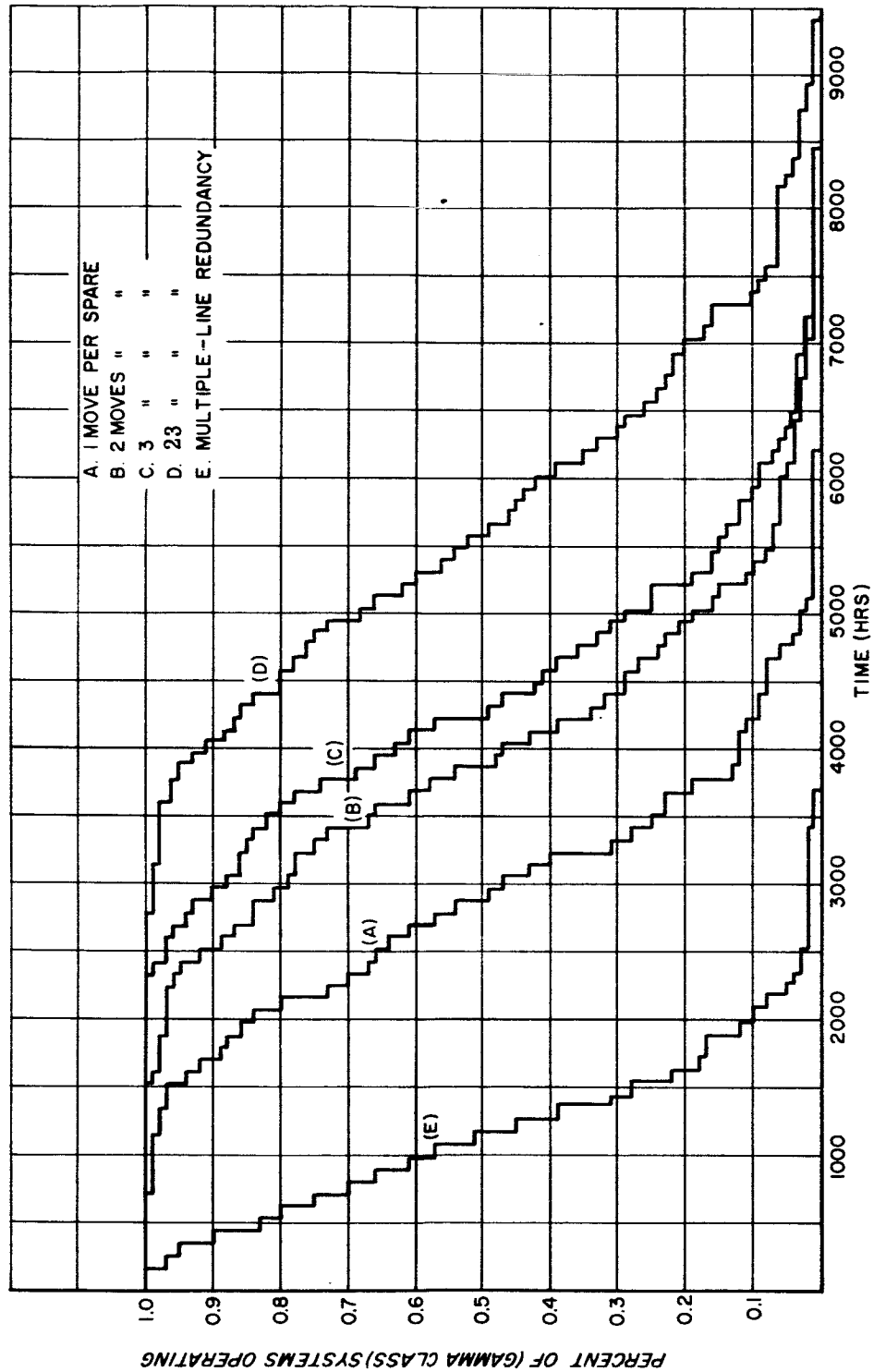


Figure 10. Percent of Systems Operating (Gamma Class 1)
Versus Time

V. SUMMARY AND CONCLUSION

Before self-repairing systems can be implemented, many feasible switching strategies must be considered in an effort to determine the most effective manner to manipulate the redundant or "spare" blocks. The extreme complexity of the reliability expressions associated with these strategies has resulted in the use of a computer simulation program for comparing the effectiveness of the strategies. Rather than proceeding to write separate programs for each strategy, a more general program has been written which employs a small number of subroutines, each of which describes an entire class of strategies. Input data determines which class subroutine is being used and which strategy in a particular class is being simulated. Although this generalized program is a great improvement over the individual program for each strategy approach, it still requires additional programming each time a new class subroutine is added. At this time, the change to a more general program, whose simulation strategy can be completely determined from input data, does not seem to merit the programming time which would be required.

The present program includes subroutines for three classes of switching strategies. Each class subroutine contains a great deal of flexibility, thereby including many individual strategies. This method facilitates easy comparison between members of a class. This comparison allows immediate elimination of many possible strategies as obviously uneconomical. For example, the flattening out of the Percent of System Failed versus Spare Mobility curves (figures 5 through 8) indicate that all possible strategies on the flat part of the curves cannot be optimum strategies.

From the results of the simulation program, curves for Percent of Systems Failed versus Spares Mobility have been plotted for the Gamma Class strategies. These curves have been referenced to that of a multiple-line majority voted system because this particular technique has been the most effective of the passive, failure masking, circuit level redundancy techniques. In all cases these curves show not only that great gains can be realized over multiple-line redundant scheme but that by far the greatest part of these gains are realized for the first few moves allowed to the spare function blocks. Beyond the range of relatively limited mobility, little or no gain in the average number of failures absorbed is realized by the additional mobility allowed to the spares. This is an encouraging result since the great majority of the gain due to self-repair can be retained without the use of an exorbitant amount of switching circuitry.

In the β and γ classes of self-repair strategies the degree of failure masking is the same as that for a multiple-line redundant system of the same order of redundancy. This is due to the fact that no "repair" is made until an ambiguity is present on the output of a

stage. This event corresponds to redundant system failure which activates the switching mechanism and the "repair" is effected. However, until the failure is "repaired" no failure masking is present, and incorrect information may be transmitted to the next stage.

The α class strategies provide additional failure masking because repairs can be initiated by the first occurrence of a failure in any stage. However, because this class implies a higher order of redundancy it cannot be compared to order-three multiple-line redundancy as the β and γ class have been.

The curves of figures 9 and 10 show a very definite gain in reliability for the self-repair strategies over multiple-line redundant systems. The curves for the Beta Class strategies show an increase in reliability for each increase in "repair" capability. Strategy β_3 yields the highest reliability but even strategy β_1 shows a significant gain over the multiple-line system. The reliability curves for the Gamma Class show essentially the same result with respect to the multiple-line case. However, investigation of the curves show that increasing the "repair" capability produces gains for the first few increases after which the magnitude of the gain diminishes. These curves tend to bear out the conclusions drawn from Percent System Failed versus Spares' Mobility curves which flattened out after a certain mobility was reached. The gains illustrated here must be considered as ideal because the switching circuitry for self-repair is here assumed to be perfectly reliable. More realistically, the gains obtainable will be a function of the switching circuitry complexity and will not be as great as shown here.

VI. FUTURE STUDIES

All of the computer simulation results discussed in this report have been based on the assumption that the switching circuitry was perfectly reliable. Efforts are now being made to determine the range of allowable failure rates which can be associated with each strategy for it to be of maximum effectiveness. These ranges are to be studied as a function of the failure rates of the associated signal processor blocks. As a result, before actual system designs are begun, information specifying the optimum switching strategy corresponding to a given signal processor failure rate should be available.

From the sample and production simulation run printouts it has become obvious that many of the spare function blocks do not experience as many switching operations as they have the capability for. When all spares are assigned a uniform mobility some reach their limit and, in doing so substantially extend the life of the system. However, in many cases when system failure has occurred, there are many spares remaining which have not been used to any great extent. In order to capitalize on this phenomenon a class of strategies γ_2 is being developed which will assign different mobilities to the spares in a stage. Class γ_2 will be simulated by a new sub-routine which is being written for the computer program. When data is available comparisons will be made between this and the other classes. Additional classes will be simulated in a similar manner as they are developed.

None of the strategies considered so far have permitted spares to return to previous locations. It is possible that removal of this restriction might add to the failure absorption capability of a system. This area certainly should be explored in this study series.

Although little has been said about the physical switching techniques to be employed, it has been tacitly assumed that the failure detection and replacement circuitry would be combined as much as possible. It has been suggested that these two phases of the repair function might profitably be separated and made almost completely independent from a circuit viewpoint. This is another area which should be given careful attention.

The Alpha class strategies have not been thoroughly investigated to determine the optimum degree of spare overlap (i. e., two sets of spares serving some of the same functional region). The information from this investigation should influence the design of new strategy classes as well as indicating the optimum strategy for the Alpha class.

VII. APPENDIX

A. CLASS α

Illustrated in figure A-1 is an α class strategy wherein each spare can "repair" failures in one row and either of two stages. Spare "1" can "repair" stages 1 or 2; "2" can "repair" 3 or 4, etc. Each spare can repair failures only in its own rows. This can be expanded such that, for example, three spares can each repair function blocks in any of ten stages or, in general, r spares for n stages. Overlapping of spares capability may help guard against "lumped" failures.

Many different strategies and system repair capabilities can be developed by simply varying r and n or by overlapping possible individual spare "repair" ranges.

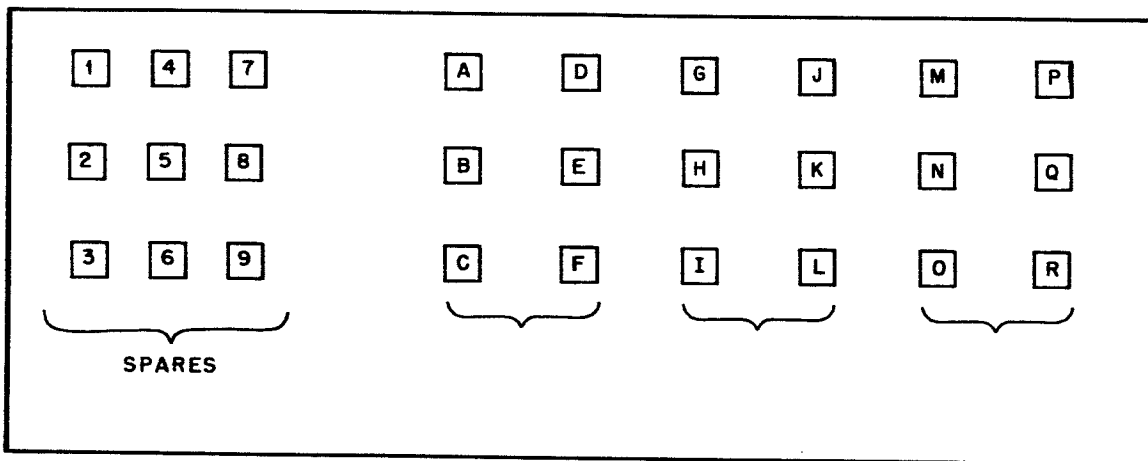


Figure A-1. Alpha Class Self-Repair

B. CLASS β

There are presently three specific strategies of β Class. The major difference between these strategies is the number of spare function blocks which can replace a given failure.

1. Class β_1 (Figure A-2)

Class β_1 allows only one "spare" for a given failure response. For example, function block "H" is given capability as a spare for stage # 4. Figure A-2a shows the system before failures occur. When one function block, J, in stage #4 fails no switching results other than the elimination of the failure. (See figure A-2b). When the second failure, say K, occurs in stage #4, function block "H" will move into stage #4 (See figure A-2c.) and resolve the ambiguity caused by the failure. After the failed block has been eliminated block "H" remains in stage #4.

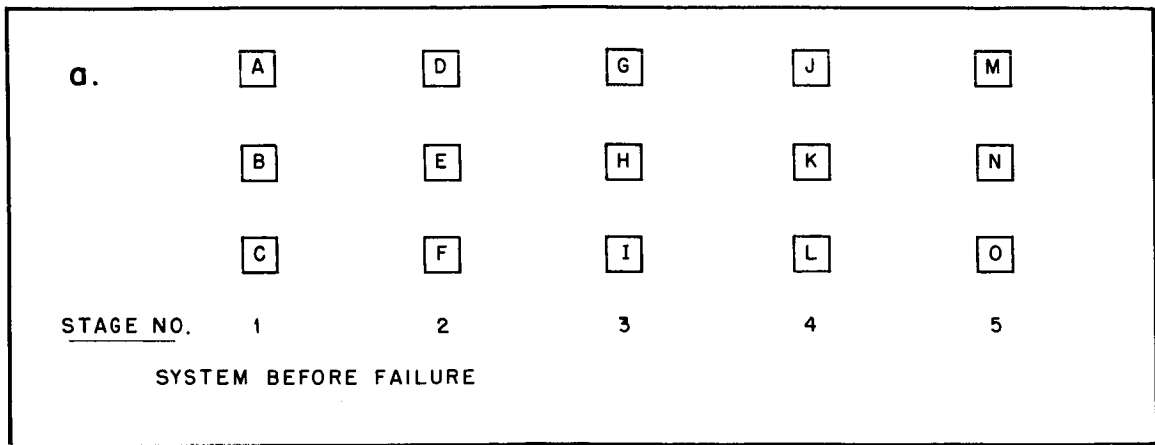


Figure A-2a. Beta Class Self-Repair

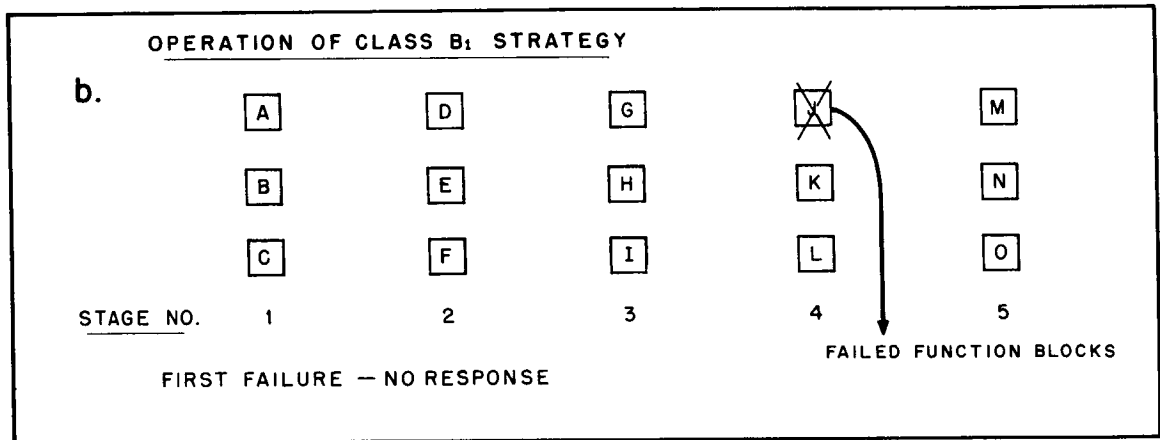


Figure A-2b. First Failure

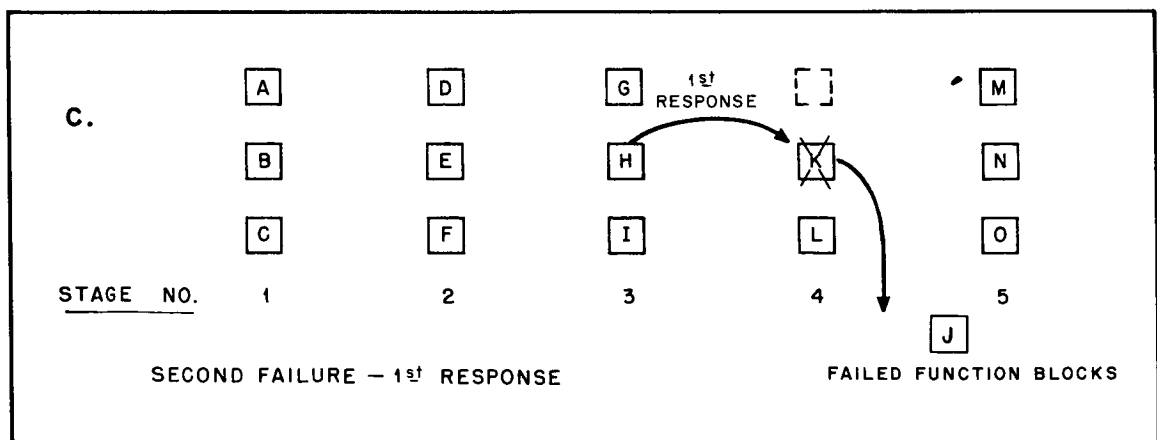


Figure A-2c. Second Failure Response

It is possible that one function block will remain working alone without system failure. For example, if function block "G" failed before "K" function block "I" will carry the load for stage 2 after "H" switches until it fails. (See figure A-3.) System failures occur when a lone operating function in a stage fails or when no spare is available to resolve an ambiguity. Failure of this system could occur when function block "E" and "G" have failed and failure of blocks "H" or "I" occurs (figure A-4), since for this strategy, block "E" is the only spare capable of "repairing" a failure in stage #3.

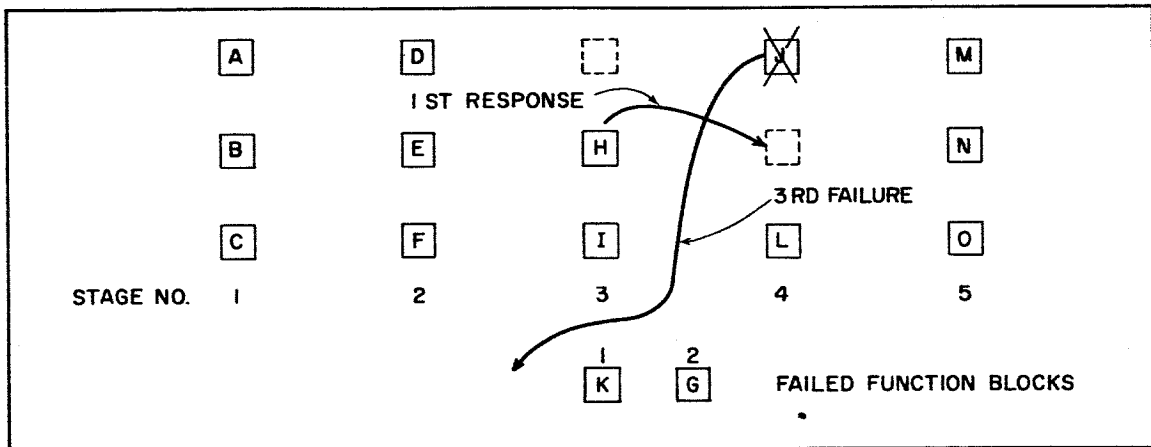


Figure A-3. Third Failure Response

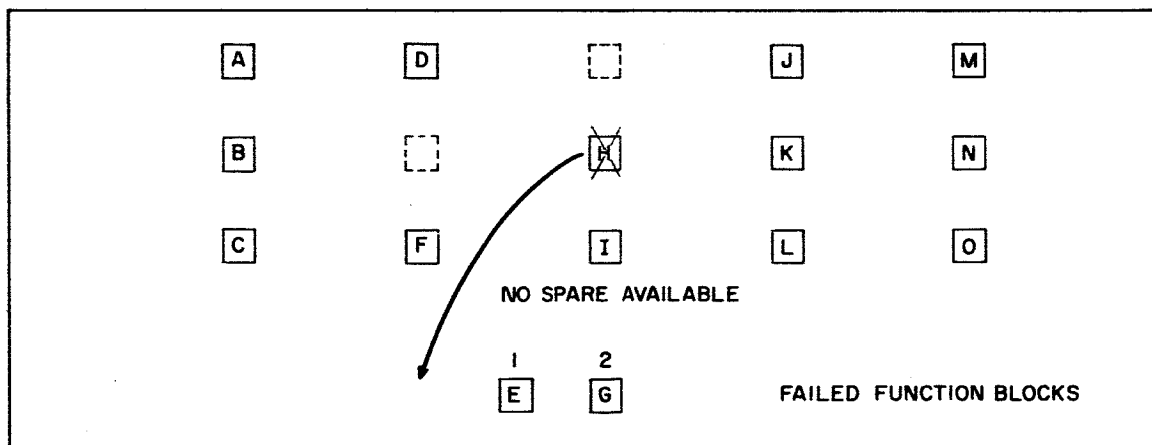


Figure A-4. Catastrophic Failure Sequence

2. Strategy β_2 (Figure A-5)

Strategy β_2 is similar to β_1 , but it allows one additional function block to replace failures in a given stage. In strategy β_2 function block "M" in addition to "H" is given the capability of replacing failed blocks in stage #4. Strategies β_1 and β_2 operate

identically through the first two failures. When the third failure in stage #4 occurs block "M", if still operative, will switch into stage #4 in the same fashion as did function block "H" in Class β_1 . This move is labeled "2 response" in figure A-5. System failure in strategy β_2 occurs in the same manner and under the same conditions as in strategy β_1 .

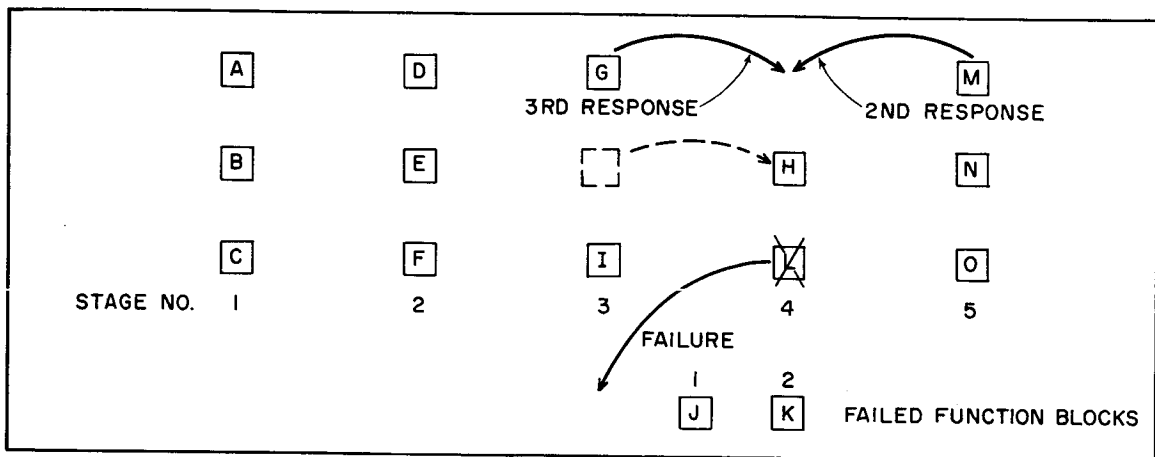


Figure A-5. Beta 2 & 3 Strategy

3. Strategy β_3 (Figure A-5)

Strategy β_3 extends the scheme one step further. Here, a third function block is allowed to move in addition to the two responses allowed to strategy β_2 . In this strategy the ability is imparted to function block "G" in stage 3 to replace failed blocks in stage #4. This is the 3rd response shown in Figure A-5. Again, failure occurs in the identical fashion to the other two strategies.

C. GAMMA (γ) CLASS

Gamma Class is divided into two parts: Class γ_1 , where all spare function blocks have the same mobility, and Class γ_2 where one spare in each stage has a greater mobility than the other.

1. Class γ_1 (Figure A-6)

As in Beta Class strategies, the first failure in a stage of a Gamma Class system evokes no response from the system. The second failure creates an ambiguity on the output of the stage. This activates the switching mechanism to switch block "H" into stage 4 thereby dissolving the ambiguity. (See Figure A-5b.) The second failed block is now identified and switched out of the system. Block "H" remains in stage 4 to detect subsequent errors. another failure occurs in stage 4, for example block "L", block "G" from stage 3 will switch into stage 4 in the same manner as did block "H". This leaves no error detecting capability in stage 2. To overcome this, block E from stage 2 switches into stage 3 to fill the void created by the switch of block "G". (See figure A-6c.)

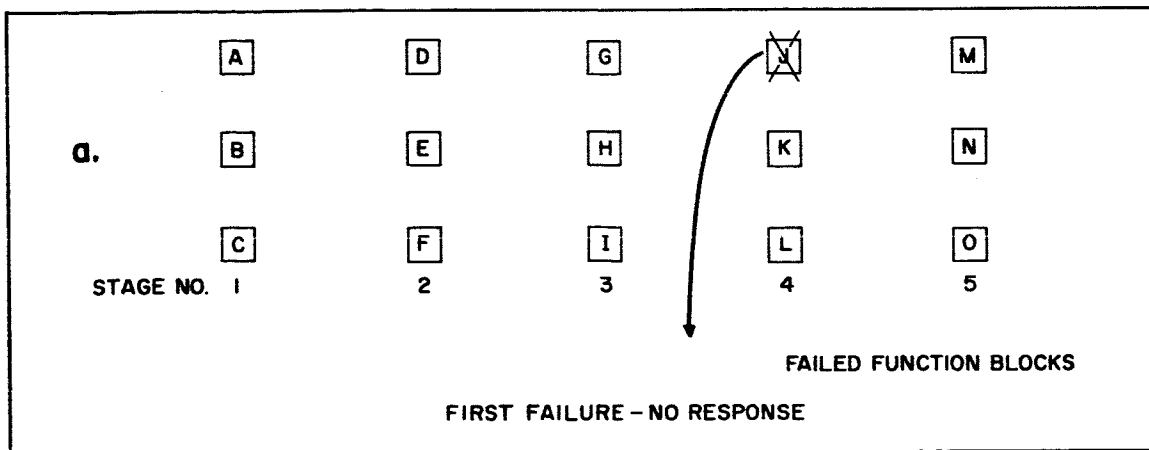


Figure A-6a. Gamma 1 Strategy - First Failure

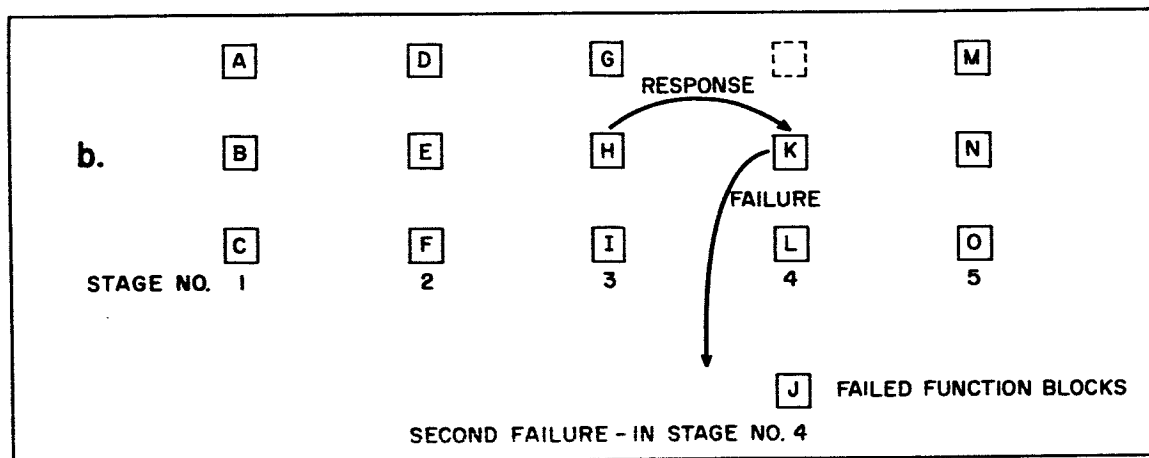


Figure A-6b. Second Failure Response

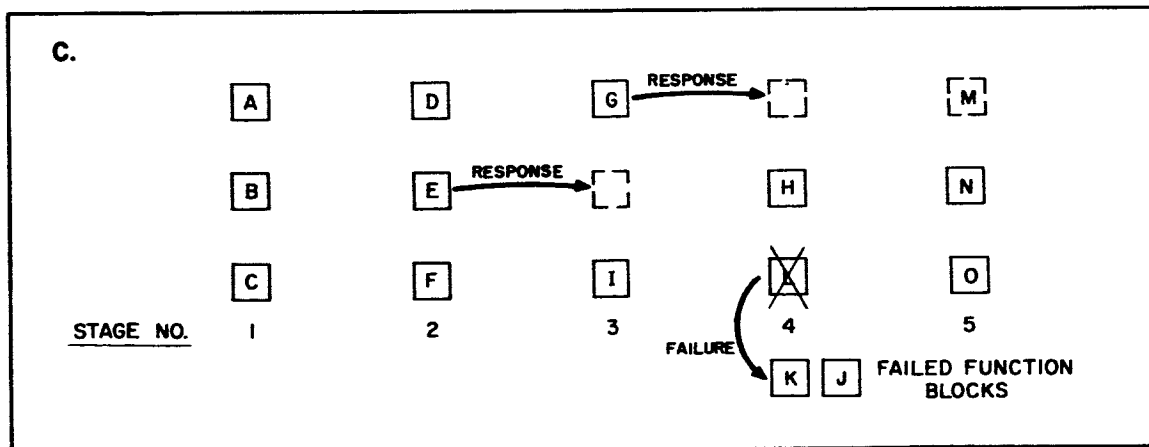


Figure A-6c. Third Failure Response

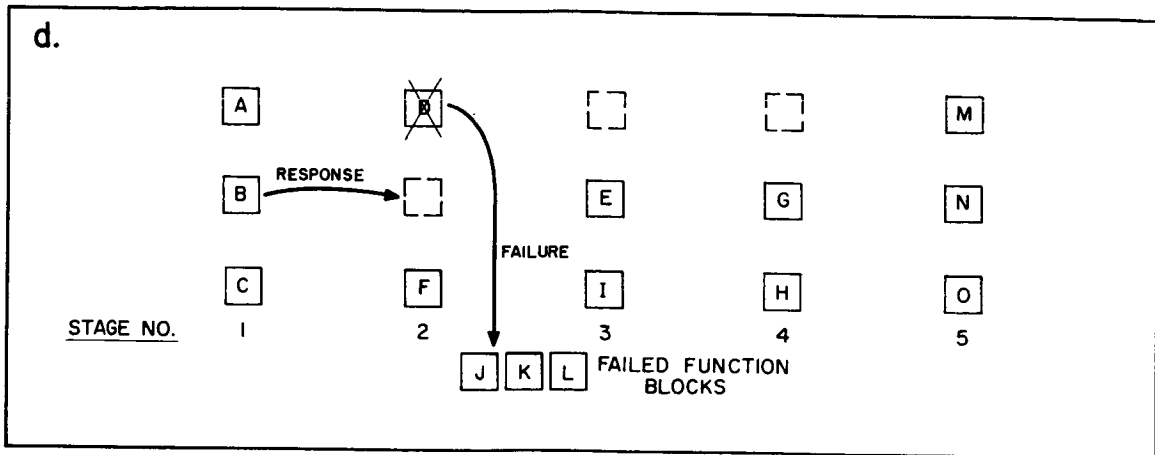


Figure A-6d. Single Block Operation

Now if a failure should occur in stage 2, block "D"; a spare function block "B", from stage 1 will switch to stage 2 and the failed block "D" will be switched from the system. (See figure A-6d.) As additional failures are sustained this process continues until a limit is reached. The end to this process can be reached in one of two ways:

1) A limit can be set for the mobility of a particular function block. In this case, once a function block has reached its limit it can no longer act as a spare for failures in the stage following it. If a critical failure occurs and all possible spares have failed or reached their limits the system fails. Voids which cannot be filled due to spares reaching their limit remain as voids but the system continues to operate until the remaining function block fails. This limit sequence is illustrated in figure A-7a. Block "A" has a

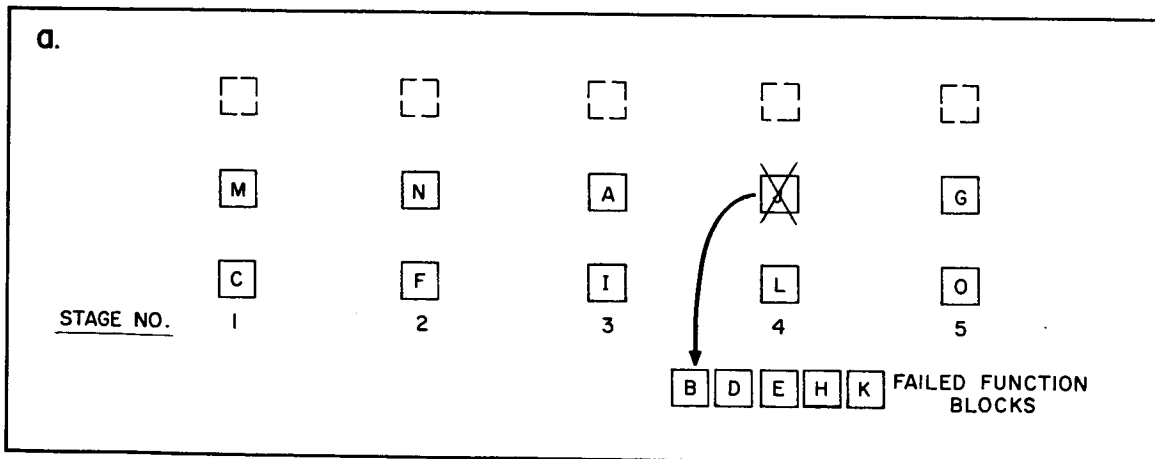


Figure A-7a. Function Block Limit

mobility of 3 and after a given failure pattern the system appears as in Figure 7a. Block "A" has reached its limit. Upon the occurrence of a critical failure in stage #4, block "A" cannot act as a spare for this stage. The ambiguity remains on the output of stage 1 and the system is considered failed. However, if the critical failure occurred in stage 2 rather than stage 4, block "M", since it hasn't reached its limit, would switch into stage 2 and resolve the ambiguity. This leaves a void in stage 1. Function block "G" cannot switch into stage 1, hence, the void remains and the system works properly as long as the remaining block in stage 1 does not fail.

2) Another failure mechanism can exist for class γ . When the system has sustained a large number of failures such that the number of remaining spares is equal to the number of stages this second mechanism case becomes effective. When an additional failure occurs, each spare function block will respond once, the initial one will resolve the ambiguity and others will fill the successive voids which appear in the immediately preceding stages. Since there is now one less spare than there are stages a void must remain somewhere in the system. If the next failure is in the stage which contains the void or that stage for which the void would have been a spare, the system goes down. For example, referring to Figure A-7b if function block "G" fails, block "D" will switch into #4 to correct for the failure. Block "A" will fill the void for block "D", block "M" for "A" and block "H" for block "M". The process stops here. There is a void in stage 5. Now failure in stage 1 or stage 5 will cause system failure. Class γ_1 , allows uniform mobility to each spare function block in the system.

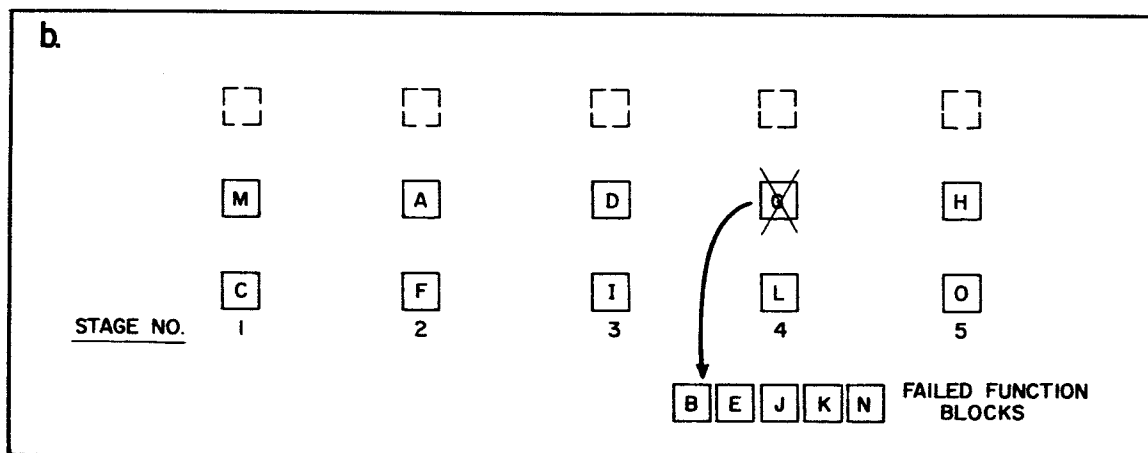


Figure A-7b. Marginal Operation

Many different strategies are contained under the heading of Class γ_1 . These differ primarily in the limit assigned to the mobility of the spare function blocks. A particular strategy may be identified by specifying "n" in the statement "n moves per spare." The value of n prescribes where a given function block will reach its limit and therefore controls the differences between the various strategies of Class γ_1 .

2. Class γ_2

Unlike the Gamma 1 Class, which assigns the same mobility to all spare function blocks, Gamma 2 Class allows the two spare function blocks to differ from one another in mobility. Figure A-8 will assist in the description of the switching processes which occur for strategy Gamma 2. The members of the top row are assigned a mobility 3, those of the middle row, a mobility 2.

The first failure in a stage will evoke no response aside from the elimination of the failed block from the system. Upon failure of the second function block in a stage (stage 4), the spare will be drawn from the next stage (stage 3). Block "G" which has the greater mobility will switch from stage 3, to stage 4. (See figure A-8a) This is the only switch which will

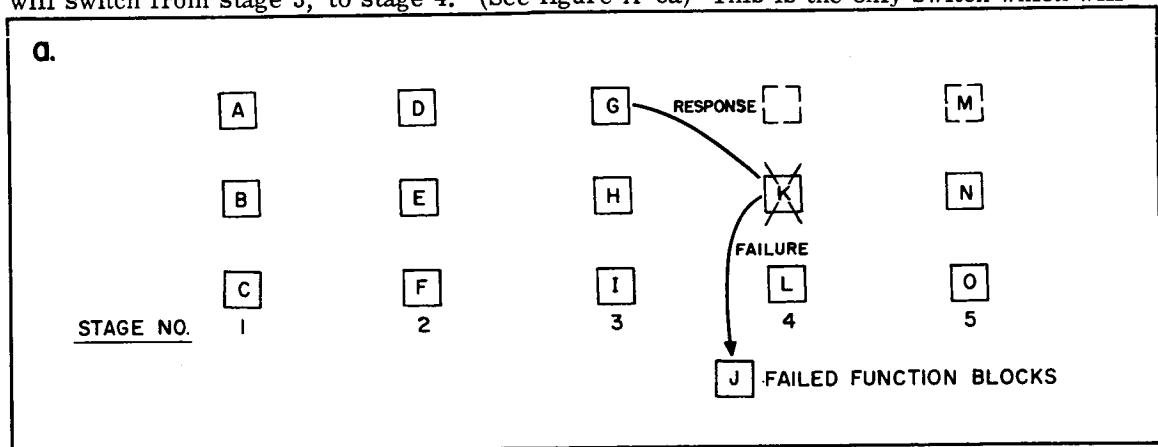


Figure A-8a. Gamma 2 Strategy - First Failure

occur. Since there are two function blocks remaining in stage 3 the void created by the switch will not be filled. The next failure occurring in stage 4 will require another spare to be switched into the stage. This spare is drawn from next stage which has a spare with high mobility and which is within range to supply the need i. e., block D from stage 2 will switch into stage 4. (See figure A-8b.) This leaves another void which is not filled and which needs not be filled. In the system described in figure A-8, the next failure in stage 4, cannot

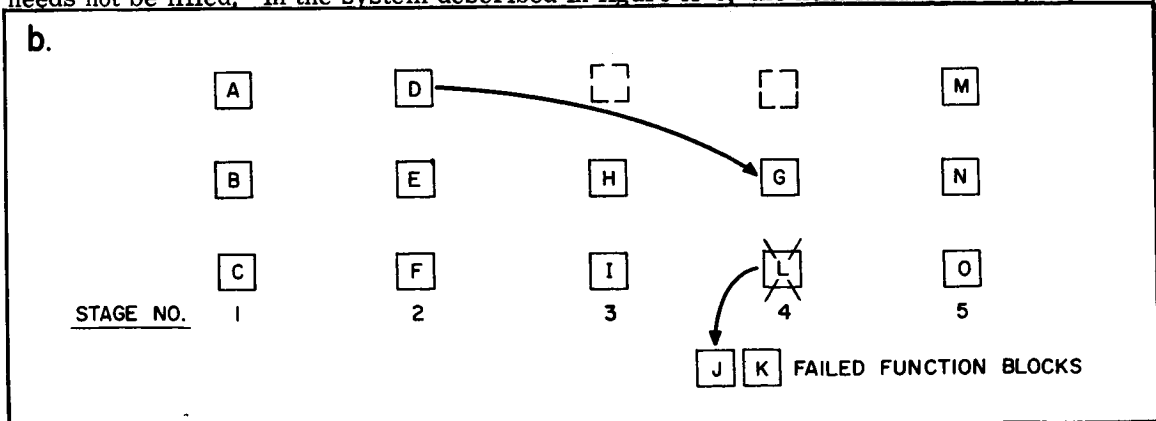


Figure A-8b. Gamma 2 Strategy

draw a high mobility spare A, because it is out of range for stage 4. In this case the lower mobility spare from stage 3 is used spare "H". This leaves a void in stage 2 which must be filled since there is only one remaining operating function block in that stage. This void is filled as though it were a failure; if a high mobility spare is available it will be switched, i. e., function block "A" will switch to stage 3. (See figure A-8c.) This process continues until either a failure occurs and no spare is available or a lone remaining function block in a stage fails. System failure occurs at this point.

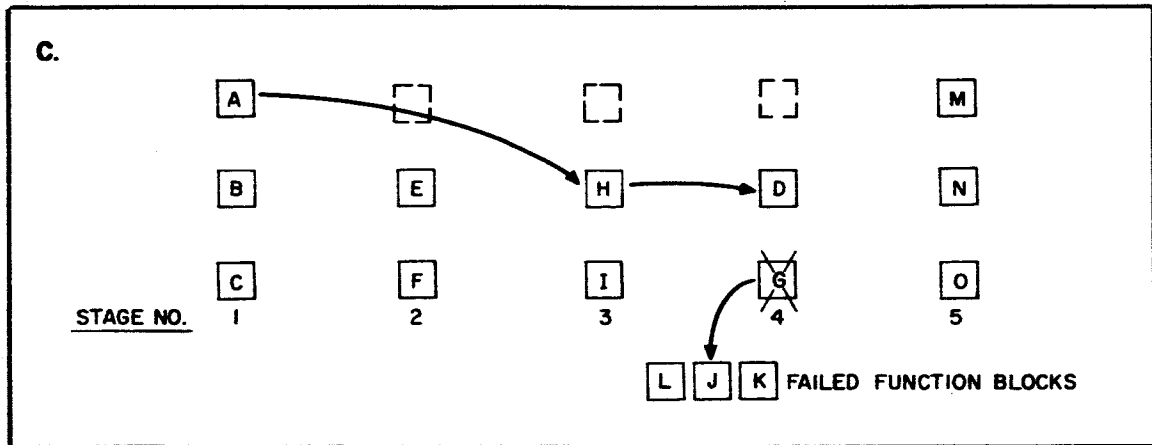


Figure A-8c. Gamma 3 Strategy - Third Failure Response